ANALYSIS

Edited by

CHRISTOPHER KIRWAN

CONTENTS

Price £1.60

ONE OF THE TRUTHS ABOUT ACTUALITY

By ALLEN HAZEN

THERE are two questions about the nature of actuality. Many writers have treated both. Some have confused them. Separated, one turns out to have a clearcut and undoubtedly correct answer. Stating it may make it easier to concentrate on the harder question.

There is the metaphysical question. Positions include extreme realism about possible worlds, with the attendant claim that there is nothing special about the actual world (Locus classicus: the chapter on foundations in David Lewis's *Counterfactuals*), as well as various kinds of actualism: positions according to which the actual world is the only fully concrete world, non-actual possible worlds being construed as states of affairs (Locus classicus: Alvin Plantinga's *The Nature of Necessity*) or propositions (Locus classicus: Robert Stalnaker, 'Possible Worlds, *Noûs*, 1976) or mathematical constructs (Locus Classicus: Quine's 'Propositional Objects' in his *Ontological Relativity and Other Essays*) or otherwise explained away.

The other question is semantical, or logical. How shall we categorize, within the framework of some semantic theory, and characterize the specific content of, words like 'actual' and its cognates? Assuming that at least some members of the family may be thought of as logical constants, what formal logic best embodies their function in our language? Here we may distinguish the *realization* theory (Plantinga, *op. cit.*), according to which actuality is a property of worlds possessed by each possible world at itself and at no other, the *indexical* theory (Lewis, *op. cit.*), according to which the (English speaking) inhabitants of each possible world call that world and no other actual, just as each (English speaking) person calls himself and no one else 'I', and the *name theory* (nobody I know; proposed as a foil to the indexical theory), according to which the function of 'the actual world' is that of a proper name for the actual world. Given the obvious relations between the various adjectives, nouns, and adverbs of the family, each of these theories provides a semantic account of the whole family, including the adjective expressing a property of possible worlds, the adjective expressing a property of possible objects, and an adverb that can be construed as a monadic propositional connective. An object is actual if it exists in the actual world, a proposition actually true if it is true at the actual world.

These different semantic answers are each compatible with any of a range of answers to the metaphysical question, but a choice between them

can be made on non-metaphysical grounds. 'Actual' and its cognates are at home in what can be broadly described as modal language. They function in sentences containing counterfactual locutions, epistemic ones, and constructions that the logician would formalize with a modal operator. Their role is to distinguish those parts of the sentence which are descriptive of the real world from those describing what might have been, what would have been if, and what is only thought to be: 'It could have turned out worse than it actually did,' 'If you had come the party would have been more boring than it actually was,' 'I thought that your yacht was longer than it actually is.'

Sometimes their function is merely contrastive. Sometimes, however, they have real logical, as well as stylistic, importance: they *insulate* some part of a sentence from its surroundings, allowing it to escape the influence of a modal, etc., locution that would ordinarily govern a part of the sentence in its position. Thus in the sentence, 'There could have been objects other than those there actually are,' the inner existential quantifier ('there are') would, on the basis of its location, ordinarily be governed by possibility operator ('could'), and so would range over the domain of the same possible world as the other existential quantifier. 'Actually,' which here may be formalized as a connective (with a good deal less distortion of the syntax than is involved in construing 'could' as one), protects this quantifier against the 'could', allowing it to range over the domain of the actual world. Or again, in the sentence 'It is logically possible for P to be the case with all actual individuals existing,' the 'all' occurs within the scope of the possibility operator, but which individuals are subsumed under it? Not those existing in some arbitrary world where P is the case, but those existing in the real world. The 'actual' allows us to quantify over *actualia* in a position where an unadorned 'all' would quantify over the *possibilia* in some, perhaps quite peculiar, possible world.

This is enough to refute the realization theory. The actuality operator it would provide would be unable to insulate subformulas of a modal sentence against the influence of modal operators. According to that theory *actually P* must be true at a world if and only if P is true at just those possible worlds which, at the given world, have the property of actuality. But on the realization theory, each world has actuality at itself alone, so this reduces to: *actually P* is true at a world just in case P is true at it. But then a sentence like *Possibly (P and actually Q)* will have the same truth conditions as *Possibly (P and Q)* rather than those we have argued it does have: those of *Q and possibly P*.

The realization theory of actuality is a mistake. It completely fails to capture, and renders inexplicable, the logical function of 'Actually'. It was, I suspect, inspired by an understandable but unfortunate way of talking current among modal logicians. Ordinarily to say of a state of

affairs that it is actual is the same as to say that it is realized, or that it obtains: ordinarily we are describing the actual world, and an actual state of affairs is simply one that is realized in the actual world. Because of this it was very easy for modal logicians, when they were talking about possible worlds and the states of affairs obtaining in them, to fall into the habit of saying a state of affairs was *actualized* in the worlds in which it obtained. But this is a special usage, and an artificial one.

That leaves us with the indexical and name theories. Each has the correct consequence that *actually* P is true at any world just in case P is true at our world. The decision between these theories must be made on other grounds.

There are possible but non-actual worlds some of whose inhabitants, speaking a language in other respects resembling English, use 'actual' and its cognates to refer to their own world. Perhaps there are also possible worlds some of whose inhabitants, speaking a language otherwise like English, use the same words to refer, not to their own world, but to ours. (I rather doubt it, but that would get us into complicated issues about the conditions under which thought or intention or verbal behaviour can single out a single entity as the referent of a term.)

If the first kind of otherworldly usage amounts to using the words with the same *meaning* we use them with, then they are indexical. What, after all, is an indexical word but one which can be used with the same meaning but different reference in different circumstances, the circumstances of its use determining its reference? If, on the other hand, it is the second kind of otherworld speaker who would be using the words with our meanings, then they are semantically more like a proper name.

There is no question in my mind about where the truth lies. Words like 'actual' are part of the fabric of the language, unlike proper names of specific individuals. One cannot be said to know their meaning if one does not use them in certain way, understand them as having a certain function. I have described this function above. One consequence of the English speaker's tacit knowledge of his language is that the inference from the negation of an instance of the Barcan Formula to 'There could have been objects other than those there actually are' is a valid one. No one using 'actually' in such a way as not to validate this inference can be said to be using it with the sense we use with it. So the inhabitants of other worlds who use 'actually' with our meaning must use it to refer to their own worlds, and the indexical rather than the name theory is correct.

Semantically, logically, the indexical theory is the truth. This does not answer the metaphysical question about the nature of possible worlds: it is a metaphysical question. Synthesis is needed, not just analysis.

Trinity College Dublin   © ALLEN HAZEN 1979

PRAGMATISM AND INTERNAL REALISM

By MICHAEL BRADIE

I. INTRODUCTION

IN *The Quest for Certainty*, John Dewey contrasts traditional epistemo-
logies with what he calls the experimental view of knowing. The root
error, he suggests, of the traditional views is that they take knowledge to
consist in a correspondence with reality ([1], 180). In 'Realism and Reason'
Putnam confesses his long support of a realist version of such a picture
which he labels 'metaphysical realism'. In 'Realism and Reason' he dis-
avows it in favour of a sanitized doctrine of 'internal realism' which he
claims to be free of metaphysical sin. Internal realism is supposedly an
empirical theory and a view that Putnam claims to have defended in two
earlier papers ([2], [3]). Putnam's case for (internal) realism is made by
a series of arguments which I shall label (1) The Success Argument, (2)
The Progress Argument, (3) The Convergence Argument, (4) The Re-
interpretation Argument and (5) The Idealistic Fallacy Argument. I sub-
mit that none of these arguments succeeds in distinguishing Putnam's
realist position from a sophisticated pragmatism which denies the
explanatory power of 'reference' and replaces 'truth' with 'warranted
assertibility (in the long run)'. Putnam notwithstanding, his disavowal of
metaphysical realism undercuts both the rationale and the basis for dis-
tinguishing realism from pragmatism.

II. THE SUCCESS ARGUMENT

The problem of success is simply: how are we to account for the
success of science? For the idealists, and here Putnam has in mind the
positivists and the operationalists, the problem is : why are our "calculi"
so useful? The "natural" realist explanation, according to Putnam,
is that our theories are successful (our calculi so useful) because nature
is *really* the way (in part) that our theories say it is.

By singling out positivism and operationalism as the idealistic
opponents to realism, Putnam faces the weakest alternatives to realism.
The pragmatic view I wish to defend as a non-realist alternative is not
committed to the view that scientific theories are mere calculi. The
pragmatic strategy is rather to challenge the notion that "the problem of
success" is capable of empirical solution. Putnam sets it up as if 'explain-
ing the success of science' were a problem as clearly defined as 'determin-
ing the speed of light'. Even if we grant, for the sake of argument, that
the problem is well defined, it seems clear that Putnam's realist answer

presupposes the rejected metaphysical realism. If internal realism is
supposed to be an empirical theory, then we should expect it to provide
empirically testable explanations. From the pragmatic point of view,
the realist explanation of the success of successful theories in virtue of
some partial correspondence with the way things really are is no more
an explanation of anything than if the positivist were to argue that
successful calculi are successful because they work.

The pragmatist solution is to challenge the way the problem is
formulated. There *is* something that needs explaining but it is the success
of *particular* scientific theories and not "the success of science". The
success of particular scientific theories is explained (if at all) by other
scientific theories. The result is an endless succession of explaining
theories. The idea that there is something more to be explained than the
success of particular theories is a mistake based on a picture that Putnam
himself ultimately rejects.

III. THE PROGRESS ARGUMENT

The problem of progress is: how to account for the progress of
science? As Putnam sees it, for the positivist, the progress of science is a
miracle. The realist has an easy answer: science progresses because our
successive theories are better and better approximations to the world as
it really is. The pragmatist has the same qualms about this "problem"
and the realist's "solution" as he has for the problem of success and its
solution. Again, Putnam's characterization of the realist solution seems to
rely on the metaphysical view that he came to reject. For the pragmatist,
as before, the catch is how 'progress' is defined. If the arguments for the
explanatory power of 'reference' can be overcome (see sections IV, V)
then the way is clear for construing progress along Kuhnian lines. Given
this construal, the realist considerations become irrelevant. Even if we
admit more historical continuity than some readings of Kuhn suggest,
it is still not clear that there is anything more than the progress of par-
ticular theories to be explained. The "progress" of these theories will be
explained, empirically, by other theories with no need to appeal to a
transcendent notion of correspondence.

IV. THE CONVERGENCE ARGUMENT

The upshot of the previous two sections is that insofar as internal
realism is taken to be an empirical theory there is an alternative position,
pragmatism, which explains "the facts" just as well. The heart of
Putnam's case against such non-realist alternatives is (1) they fail to

account for what Putnam, following Boyd, calls 'convergence', and (2) they involve a reinterpretation of scientific knowledge which is belied by scientific practice. We consider each point in turn.

For Putnam's purposes, 'convergence' boils down to two 'facts':

Fact 1: Scientists try to preserve the mechanisms of established theories.

Fact 2: This strategy [*sometimes*] leads to important discoveries.

Putnam's empirical theory of realism (for which he expresses indebtedness to Richard Boyd) consists of two theses:

BR 1: Terms in a mature science typically refer.

BR 2: The laws of a theory belonging to a mature science are typically approximately true.

The problem of convergence consists in explaining the two facts. Putnam puts forward two arguments which show that Boyd's realism accounts for both facts ([2], p. 180). They are:

Argument 1: Scientists accept BR 1 and BR 2.

explains

Fact 1.

Argument 2: BR 1 and BR 2 are true.

explains

Fact 2.

Putnam claims that replacing 'are ... true' in BR 2 by an "operationalist" substitute, 'are simple and lead to true predictions', fails to preserve the validity of Argument 1 and Argument 2. Argument 1 fails, presumably, because with the substitution there would be no incentive to try to preserve the mechanisms of earlier theories. Argument 2 fails, presumably, because if there is no correspondence, there is no reason to expect the strategy to lead to important discoveries.

The pragmatist has three points in response. First, the qualifier '*sometimes*' in Fact 2 (which Putnam omits) is crucial. Since the strategy sometimes *fails*, the significance of Fact 2 is suspect. It is a fact of history that some of our theories are more fruitful than others. Why those that are fruitful are so, and those that are not are not, is to be explained, if at all, by further theories. Second, although Putnam says he is going to show how an instrumentalistic view fails to explain the behaviour of scientists *or* the success of their strategies, he winds up arguing only for the weaker thesis that without the notions of 'truth' and 'reference', the class of potential successor theories is much wider and the chances of successfully choosing a suitable successor go down accordingly ([2], p. 181). This might be so, if we were still wedded to the metaphysical correspondence picture. But, even so, it is not clearly the case. Without the metaphysical picture, the success of certain strategies is not

miraculous but, again, a matter for some future theory to account for. Third, Fact 1, in so far as it is a fact, can be given a sociological explanation in terms of the "institutional practices of the day". Scientists act the way they do because, in part, of the way they are trained, partly, no doubt, because such practices often do work, and partly because of the immense intellectual effort it would cost to completely revamp existing models and mechanisms at the slightest whim. Perhaps, in a computer dominated society of tomorrow, when alternative models and mechanisms are a dime a dozen, such conservative practices will appear to be among the relics of a bygone age.

V. THE REINTERPRETATION ARGUMENT

What if there were no convergence of scientific knowledge? Putnam suggests that (almost) (1) if convergence is denied, then referentiality fails (i.e., if there is no convergence then *no* theories refer) [Pragmatist: so much the worse for "the very idea of reference"]; (2) if referentiality fails, then the *truth* of our theories fails ([2], p. 184). What actually happens, Putnam argues, is that the pragmatic treatment of 'truth' and 'reference' boils down to a reinterpretation of the connectives of classical logic, and *not* a rejection of the "formal" structure of 'true' and 'refers'.[1] We can concede Putnam's point that, at best, the realist/idealist dispute (as he sees it) comes down to alternative interpretations for a shared formal structure. Putnam's case then rests on the need to show that interpreting the logical connectives realistically is preferable to interpreting them non-realistically. Putnam attempts just this. He first establishes what it is to interpret the connectives realistically and then argues that this interpretation is preferable to all others ([2], sec. 6). Putnam's case hinges on the realist being able to accept certain types of statements which the non-realist is forced to reject. Putnam cites as typical the two statements:

(A) Venus could have carbon dioxide in its atmosphere even if it didn't follow from our [current] theory (... plus the set of "true observation sentences") ...

(B) A statement can be true even though it doesn't follow from our [current] theory (... plus the set of "true observation sentences"). ([2], p. 192).

The realist accepts (A) and (B). (A) entails (B). (A) is itself a statement

[1] Thus, even if we were to replace 'is true' with 'is warrantedly assertible', it is possible to define both in a Tarskian manner such that each satisfies a version of Tarski's Criterion T ([2], sec. 3). This argument depends upon some controversial means of re-reading the "classical" logical connectives against which Kripke (in lectures) has raised some compelling arguments.

which any scientist would accept, Putnam suggests, whatever his philosophical persuasion. (B) is the heart of realism. How, then, is realism to be avoided? By one of two strategems:

(IS 1) Taking (A) at other than face value by reading the predicates "idealistically". This leads to phenomenalism and logical empiricism, which are rejected programs. Hence, IS 1 is unacceptable.

(IS 2) Since the predicates cannot be reread, the connectives must be reread and we 'save the bulk of extensional scientific theory, and save the formal part of our theories of reference and truth, *at the cost of giving up* (*A*) *and* (*B*)'. ([2], 193; my italics).

To block IS 2, Putnam appeals to Boyd's concept of convergence, which shows that 'this sort of anti-realism, "cultural relativist" anti-realism, is bankrupt'. ([2], 194). Given the justice of our earlier remarks on Boyd's program, this celebration may be premature.

The key question centres on the supposed cost Putnam sees for going the anti-realist route. Need the anti-realist give up (A) and (B)? I argue below that the pragmatist need not. If not, then Putman's case for realism collapses. If the pragmatist can hold on to (A) and (B), and do so without invoking correspondence or reference, then the realist is put on the defensive, at least to the extent he wishes to retain these (pragmatically speaking) "idle wheels".

Putnam's confidence that the anti-realist will be forced to reject (A) and (B) is, I believe, misplaced. He is confident of this only, seemingly, because of his identification of the anti-realist as someone akin to 'an Intuitionist or quasi-intuitionist'. ([2] p. 193). On this reading, 'true' is supposed to be replaced by 'is provable on the basis of our theory' and, hence, (B) now reads

(B') A statement may be provable on the basis of our theory even though it does not follow from (is not provable on the basis of) our theory.

Such a reading renders (B) patently unacceptable. Since (B) follows from (A). (A) is rendered unacceptable as well.

If this is the rationale for Putnam's confidence, and I see no other, it will not do the job. There is another non-realist reading of (B) which is closer to the spirit of (B) and does not involve a patent absurdity. Instead of replacing 'true' by 'provable on the basis of our theory', we replace it by 'provable (in the long run)'. (B) now reads

(B") A statement may be provable (in the long run) even though it doesn't follow from (is not provable on the basis of) our [current] theory.

Surely this is acceptable and captures the essence of the pragmatic view,

that future theories may correct and supersede earlier ones (see section VI). But, an objector might ask, don't we still need the notion of 'really true'? This amounts, the pragmatist claims, to no more than the demand for a theory which will never be superseded, a demand which, for all we know, may one day be realized, but hardly one that science at its "face value" commits us to.

VI. The Idealistic Fallacy Argument

In [3], Putnam sketches what he calls a 'causal theory of reliability' which is to account for how we are successful knowers and communicators. He argues that the causal picture (based on the best scientific knowledge) cannot be easily accommodated by non-realist positions. He suggests that the modal statements present the gravest difficulties. He then puts forward what he calls the 'idealistic fallacy argument.' ([3], p. 16).

IF: 1. On the causal theory of reliability, there is room for error.

2. 'The rug is green' might be warrantedly assertible even though the rug is not green.

But 3. On the pragmatist view, 'warrantedly assertible' is interchangeable with 'true.'

Then 4. 'The rug is green' might be true even though the rug is not green.

But this is impossible.

5. 'Warrantedly assertible' is NOT interchangeable with 'true.'

6. Pragmatism (and idealism in general) is untenable.

This difficulty seems easily overcome, for what the pragmatist would allow is that

'The rug is green' might be warrantedly assertible *given all we know* even though the rug is not green.

Thus, once we realize that what replaces 'true' is not 'warrantedly assertible on the basis of our present information' but rather 'warrantedly assertible in the long run', Putnam's argument loses its force. It shows, perhaps, that truth can't be warranted assertibility (in the short run), but it doesn't show that 'warranted assertibility (in the long run)' doesn't do everything that 'truth' does. In fact, this is just the position Putnam himself comes to in [4] ([4], pp. 3 ff.).

VII. Conclusion

I have argued that Putnam's disavowal of metaphysical realism has
cut the ground from under his attempt to distinguish internal realism
from any non-realist alternatives. Five key arguments in Putnam's defence
of internal realism have been shown to be inadequate to distinguish inter-
nal realism from pragmatism. The result is that a version of pragmatism
which contrasts warranted assertibility (in the long run) with warranted
assertibility (in the short run) seems capable of doing everything that
Putnam's internal realism does but without invoking the suspect notions
of 'correspondence' and 'reference'.[2]

Bowling Green State University   © Michael Bradie 1979

REFERENCES

[1] Dewey, John.: *The Quest for Certainty*, Putnam, 1960.
[2] Putnam, Hilary.: 'What is Realism?,' in *Proceedings of the Aristotelian Society*, NS, vol.
        LXXVI, 1976.
[3] —— 'Reference and Understanding,' in H. Putnam, *Meaning and the Moral Sciences*
        Routledge and Kegan Paul, 1978.
[4] —— 'Realism and Reason', in Proceedings of the APA, 1976–77, vol. 50, No. 6.

THALBERG ON THE IRREDUCIBILITY
OF EVENTS

By RICHARD H. FELDMAN *and* EDWARD WIERENGA

IN a recent paper Irving Thalberg argues for the irreducibility of events (ANALYSIS, 38.1, pp. 1–9). He never says exactly what it is to reduce events, but it seems clear that there are at least two quite different projects that event theorists might undertake, each of which could appropriately be termed 'reductive'. In evaluating Thalberg's arguments it is especially important to distinguish these two kinds of reduction since of the two theories he considers one is reductive in one way and the only reductive version of the other is reductive in the second way.

The first kind of reductive theory of events is an event-language reduction. A proponent of this kind of theory attempts to show that all true sentences that seem to contain references to events can be paraphrased into sentences that clearly do not contain references to events. There are sentences that we know to be true that seem to refer to such concrete events as Napoleon's defeat at Waterloo or Socrates' death. Such sentences might report the causes and effects of such events, our attitudes about them, or, perhaps, the locations in which they occur. An event-language reduction theory would provide rules for paraphrasing these sentences into sentences about other entities. So, according to these theories, event talk can be reduced to non-event talk. Defenders of event-language reduction theories hold that there are no such things as concrete events, where concrete events are understood to be non-repeatable, contingent particulars.

The second kind of reductive theory of events is an event-reduction theory. Theories of this sort imply that events do exist, but they are just a species of some other, possibly more familiar, kind of entity. For example, according to a theory to be discussed later, events are sets with certain sorts of members. Events are thus reduced to sets of the appropriate kinds.

Roderick Chisholm has defended the most well-developed versions of an event-language reduction (1970, 1976). Chisholm recognizes the existence of "generic events"—eternal, non-contingent entities which are in many cases repeatable. He believes that such events are states of affairs. Chisholm asks,

> Must we say ... that in addition to [generic] events there are such things as the *particular occurrences* of events? Or can we reduce talk that is ostensibly about the particular occurrences of events to talk that refers just to [generic] events, attributes, and individual things? (1970, pp. 16–17).

Chisholm goes on to argue that talk of concrete events (and generic events as well) may be reduced to talk of states of affairs. So it is

clear that Chisholm holds an event-language reduction theory, despite Thalberg's claim that Chisholm might not hold a reductive theory at all (p. 3).

Chisholm's view is that all true sentences that are ostensibly about concrete events may be paraphrased into sentences about states of affairs. States of affairs are things much like the propositions of other theorists. They exist necessarily. They are individuated rather finely: Carter's arriving and the President's arriving are diverse states of affairs despite the identity of Carter and the President. Analogous to the view that every proposition is either true or false, every state of affairs either obtains or does not obtain at any given time. For example, Brutus' being Dictator is a state of affairs that exists, although it never obtained. Caesar's being Dictator did obtain for some years, but at all other times it existed without obtaining.

The details of Chisholm's paraphrases need not concern us here. Some are quite complicated and it may be that not all of them are entirely satisfactory. (For some objections to the details of the earlier version of Chisholm's theory see Wierenga, 1976.) What is important is that true sentences ostensibly containing reference to concrete events are eliminated and replaced by sentences in which various states of affairs are said to obtain or not obtain.

Obviously, our understanding of Chisholm's theory depends in large part upon our understanding locutions basic to the theory such as 'State of affairs *s* obtains'. Thalberg professes not to understand, and this is his first objection to Chisholm's theory. He points out that in English 'obtains' is a transitive verb meaning 'gets hold of by effort'. Based on this observation he brings out two difficulties for Chisholm's basic locutions: first, inanimate things like states of affairs cannot have goals or make efforts, so they cannot obtain anything; second, Chisholm's basic sentences are grammatically incomplete, lacking a direct object phrase (p. 4).

These complaints seem to us to be based in part upon a refusal to understand. Although 'obtains' does have the meaning Thalberg describes, it also is used as an intransitive verb meaning 'is in force' or 'is established', as in 'Peace obtains'. It is something like this second meaning that Chisholm has in mind in his basic sentences.

'Obtains' is an undefined term in Chisholm's theory. It expresses a property some but not all states of affairs have. Chisholm suggests that this property can also be expressed by 'occurs', 'takes place', and 'is actual' (1971, pp. 39–40). States of affairs that obtain can be said to be 'realized or manifested' (1976, p. 120). Chisholm also states a number of principles or axioms pertaining to his basic sentences (1976, pp. 117–120). So it seems that he has provided an adequate explanation of his reasonably clear primitive predicate.

Thalberg thinks that Chisholm's theory faces a 'deeper difficulty' than the puzzles about 'obtains'. He writes:

> ... to what extent has this kind of theory elucidated our notion of an occurrence? What metaphysical insights does it give us into the nature of events? Suppose we agree that there are 'two sorts' of states of affairs, by analogy with the way propositions are either true or false. Next we interpret any of the verbs we have chosen to mark this dichotomy— 'obtains', 'takes place', 'is actual', 'exists' and their antonyms—as literally as possible. We seem to be describing something that happens, or perhaps fails to happen, to some quasi-propositional entity. This sounds suspiciously event-like. We appear to have elucidated events generally ... by reference to one special kind of happening which only features states of affairs (p. 4).

He later says that this theory 'leaves us with an event: the "occurring" or "taking place" of a state of affairs' (p. 5).

Thalberg's objection seems to be that the theory tells us that for a concrete event to occur is for a state of affairs to obtain. But if a state of affairs obtains, then there is such a concrete event as the obtaining of the state of affairs. So we are left with concrete events.

The problem with this objection is that we have been given no reason to think that if a state of affairs obtains then there is a concrete event which is the occurrence of that state of affairs. Chisholm's theory does not appeal to happenings which feature states of affairs. Rather, it says that there are no such things. True sentences purporting to involve reference to particular occurrences or happenings are provided with paraphrases that are about states of affairs and their properties. So Thalberg's main objection to Chisholm seems to be incorrect.

In the first two sentences of the passage just quoted, Thalberg criticizes Chisholm's theory by suggesting that it gives us no insight into the nature of concrete events. But clearly Chisholm doesn't tell us what concrete events are because, according to his theory, there aren't any. Thalberg's criticism seems to be based on construing Chisholm's theory as an event reduction rather than as an event-language reduction.

It is possible to object to Chisholm's theory on a variety of grounds. One might criticize some of the specific paraphrases he proposes. One might contend that the entities he appeals to—states of affairs—are obscure. One might argue that states of affairs language can be reduced to event language, and that effecting such a reduction would be preferable to Chisholm's approach. We do not intend to endorse any of these objections or to defend Chisholm from them. Our contention is that Thalberg is mistaken in claiming that if the theory is meant to be reductive, which it is, then it is a failure. He has provided no reason to think that Chisholm has not given an adequate reduction of event language to states of affairs language.

The second theory of events Thalberg criticizes is the property-exemplification theory defended by Goldman (1970, 1971) and Kim (1966, 1973, 1976). According to one version of this theory an event is the exemplification of a property by a substance at a time. For example, the concrete event which is Socrates drinking hemlock at time t is the exemplification by Socrates of the property *drinking hemlock* at time t. This can be represented by the formula '[Socrates, drinking hemlock, t]'.

As Kim states his view, it is not reductive. He 'does not attempt to show that events are in some eliminative sense "reducible" to substances, properties, and times' (Kim, 1976, p. 163). Rather, he says they are 'structured complexes' whose constituents are objects, properties, and times. The event itself is the object's having or exemplifying the property at the time. Thus, as Thalberg says, this theory is not reductive.

Even though this property-exemplification theory is not reductive, it might be said that it clarifies our thinking about concrete events by showing that all events are of one basic kind: the having of a property by a substance at a time. Thalberg protests that there is no clarification here:

> In what sense does an episode of 'exemplifying' underlie all other events? ...How would it 'lay bare the nature, or ontological status' of such events as the 1967 Chicago blizzard, and Socrates' ingurgitation of hemlock, if we learn that these are really cases of an object exemplifying a property? (p. 8).

It's not clear that Kim or Goldman ever said that exemplifications 'underlie' all other events, but it is clear that we do learn something if their theory is correct. We learn the identity conditions for events: events are identical provided they are exemplifications of the same property by the same substance at the same time. In the opening paragraph of his paper Thalberg points out that a better understanding of events would help resolve a number of philosophical issues, including the mind-brain identity problem and the relation of human actions to bodily movements. Armed with what we learn from Kim's and Goldman's theory, we can begin to try to resolve these issues.

In a recent paper Kim describes, but does not explicitly endorse or reject, an event-reductionist variation on his theory. He says that his

> theory can easily be developed along a different line: dispense with the existence condition and define the *predicate* 'is an event' over ordered triples of substances, properties, and times. An ordered triple $\langle x, P, t \rangle$ would be an event just in case the substance x has the property P at time t. The existence of the triple would be guaranteed, by the principles of set theory, provided x, P, and t exist, whether or not x has P at t (1976, p. 162).

It would be a mistake to object to this theory by claiming that a given event is not a triple on the grounds that the triple, but not the event, could have existed even though the constitutive substance did not exemplify the constitutive property at the constitutive time. Under those

circumstances, the event *could* have existed, but would not have been an event—it would have been a mere triple. These mere triples can be taken to be possible events. Allowing for possible events might be an advantage of the theory.

This variation of the property-exemplification theory is reductionist in the second of the two ways discussed above. It reduces concrete events to set-theoretic entities whose members are substances, properties, and times. Thus, on this view, it must be admitted that there are events, but they are nothing "over and above" sets of the appropriate kind.

Some philosophers might find it counter-intuitive to regard events as sets and some might not like the sets, properties, and times to which this theory is committed. But these misgivings do not constitute objections to the theory. Until some objections are found, this theory stands as a possible event-reduction theory.

Thalberg does raise one objection that may be construed as an objection to this variation of the property-exemplification theory. He argues that property-exemplification theorists have not avoided reference to events (as if they were trying to) because in referring to a time they somehow make reference to events. This is because

> Every dating system which is used to specify moments and intervals of time does appear to make explicit reference to such happenings as the periodic appearance of sunlight, the change of direction of the shadows it casts, and its disappearance (p. 7).

It is difficult to know what to say about this objection. It is true that our dating systems contain expressions that make reference to events. On the view under discussion, these events are ordered triples of the appropriate kind. The fact that we use such references in our talk of times seems to be of no great importance here—it in no way conflicts with the claim that events are triples of substances, properties, and times.

It is possible that in making this objection Thalberg has in mind the view of some philosophers that times can be reduced to events (see Russell, 1927). Since it would be wrong simultaneously to reduce times to events and events to times, it is obvious that a defender of this theory of events should not accept this reductionist view of time. He should say that times are irreducible, or else he should reduce times to things other than events. Unless there is some compelling reason to think that times ought to be reduced to events, this is no objection to the theory.

Perhaps Thalberg's point is that we are unable to refer to times without at some time fixing the reference of our time words by using words that refer to events. Thus if e_1 and e_2 are events, we might say that time 1 is the time at which e_1 occurred and that a unit of time is the duration of e_2. As a result, all our references to times depend upon prior references to events.

Even if this is true, it is difficult to see why it is a problem for the event reduction theory described here. However, it may be a problem for an event-language reduction such as Chisholm's most recent theory, since references to times are made in that theory (although they are not in his earlier theory). If Thalberg's objection is correct, Chisholm must make reference to events in order to make reference to times. So he is not able to eliminate all apparent references to events: he must retain those used in establishing the referents of his time words.

However, it is not at all clear that Thalberg's objection is correct. We might fix the referent of the expression 'Time 1' by saying that it shall refer to now (the present moment). A short time later we might establish the referent of 'Time 2' in a similar way. A unit of time would then be the interval between Time 1 and Time 2. Such a system might be awkward to use unless a unit of time coincided with some regular and natural occurrence. Although such a connection between time words and regular events is a practical convenience, it does not seem to be a logical requirement for references to times. So it is implausible to think that we can't refer to times without referring to events.

We conclude that Thalberg has given no good reason to think that Chisholm's theory is not an effective event-language reduction or that a variation on the Kim-Goldman property exemplification theory is not an adequate event-reduction theory. Hence, contrary to his contentions, his reasoning does not support the conclusion that no reductive theory of events is successful.[1]

University of Rochester   © RICHARD H. FELDMAN *and*
                                                                EDWARD WIERENGA 1979

REFERENCES

Chisholm, R.M. (1970) 'Events and Propositions', *Nous*, IV, pp. 15–24.
—— (1971) 'On the Logic of Intentional Action', in R. Binkley *et al.* (eds.), *Agent, Action and Reason*. Toronto: Univ. Press, pp. 38–69.
—— (1976) *Person and Object*. London: George Allen & Unwin.
Goldman, A. (1970) *A Theory of Human Action*. Englewood Cliffs: Prentice Hall.
—— (1971) 'The Individuation of Action', *The Journal of Philosophy*, LXVIII, pp. 761–74.
Kim, J. (1966) 'On the Psycho-Physical Identity Theory', *American Philosophical Quarterly*, III, pp. 227–35.
—— (1973) 'Causation, Nomic Subsumption, and the Concept of Event', *The Journal of Philosophy*, LXX, pp. 217–36.
—— (1976) 'Events as Property Exemplifications', in M. Brand and D. Walton (eds.), *Action Theory*. Dordrecht: Reidel, pp. 159–77.
Russell, B. (1927) *Outline of Philosophy*. London: George Allen & Unwin.
Wierenga, E. (1976) 'Chisholm on States of Affairs', *Australasian Journal of Philosophy*, 54, pp. 148–152.

ASSERTIONS, PLAIN AND COLOURED

By ARNOLD SPECTOR

IT is sometimes instructive to take certain expressions and see what modifiers or transformations they refuse. Consider, for instance, the *if . . . then* construction. If '*p*' and '*q*' stand for assertable sentences (that is, sentences which can stand on their own and have a truth-value on their own), then

> (1) If *p*, then apparently (no doubt, hopefully, regrettably, fortunately) *q*

as well as

> (2) Apparently (no doubt, hopefully, regrettably, fortunately) if *p*, then *q*

are perfectly acceptable whereas

> (3) If apparently (no doubt, hopefully, regrettably, fortunately) *p*, then *q*

is difficult to interpret, or at least sounds strange. In the case of the *either . . . or* construction both disjuncts tend to reject many modifiers:

> (4) Either he's apparently (no doubt, etc.) highly paid or he's of independent means

sounds just as awkward as

> (5) Either he's highly paid or apparently (no doubt, etc.) he's of independent means.

In a similar fashion when sentences containing certain modifying expressions are put into the interrogative form we are left with a feeling of unease. When the following sentences

> (6) Jones is conceivably the best candidate
> (7) Jones is no doubt the best candidate
> (8) Jones is hopefully the best candidate
> (9) Jones is deplorably the best candidate

are transformed into questions the results have an air of strangeness:

> (10) Is Jones no doubt the best candidate?
> (11) Is Jones hopefully the best candidate?

(For the sake of exposition, I assume that the questions (10) and (11) are uttered with normal intonation—a rising contour—which shows that they are being used to *initiate* a piece of dialogue and are requests for information and are not spoken with any other pattern of intonation such as, say, that which is used to raise a question about something

which somebody else has already said, in which case the 'is' would be stressed and have a falling contour.)

I should like to suggest that the problems contained in (10) and (11) are related to the problems that spring from (3); they are all to do with assertion. I admit that my examples depend quite a lot on the specific modifier, not on wholly general principles; nevertheless, I think they point to some interesting tendencies in the conventions of making assertions.

Examples (6)—(9) may be taken to possess the same propositional content, or thought, to use Frege's jargon, namely that Jones is the best candidate. (Even if the force of a certain operator shows that the speaker asserts the negation, or partial negation, of a sentence, the thought itself remains unaltered.) Examples (6) and (9) contain expressions of assertoric force and (7) and (8) contain expressions of assertoric colour. It seems that there is in human discourse a spectrum of assertoric force and assertoric colour. It may be recalled that for Frege a sentence contained two strands of meaning; one was its sense, that by which its truth or falsity was determined, and the other strand was its colour, which was everything else that was conveyed by the sentence. I wish here to speak of the force and colour of *asserting*: the truth of a thought can be asserted with an abundance or a want of vigour. Further, the speaker can paint his assertion with various colours, the colour of an assertion being anything else that is conveyed by its expression other than the energy with which its truth is asserted.

There is, connected with the issues underlying this discussion, a debate on whether to regard assertion as something psychological, as a manifestation of a mental state (as did Frege, Russell and the youthful Wittgenstein) or to treat it as a publicly observable operation performed according to certain linguistic rules on the expressions of a thought (as does Dummett, say). The latter account of assertion as an exterior act is more congenial to the notion of assertoric force than to that of assertoric colour. (A man may be said to kick a football tentatively or with all his strength but may not be said literally to kick a football regrettably or fortunately.) But even for assertoric force, the account of assertion as a conventional action is problematic. For there seems to be a distinction that can be drawn between asserting something without much confidence and guardedly asserting it; there seems to be a distinction between degrees of confidence and degrees of assertion. A man who has left his spectacles at home may assert without much confidence of a figure in the distance, 'That *may* be my wife'; a man to whom report has been brought of some misdemeanour on the part of his wife may assert guardedly, 'That *may* be my wife'.

Having admitted these scruples, I intend to suspend them so far as this discussion is concerned and I will assume that assertion may be

characterised as a linguistic act and not a manifestation of a mental state.

Moreover, I am going to suggest that the unease caused by such examples as (3), (10) and (11) is grounded on a feeling for the conventions of asserting.

I propose that sentences (6)—(9) possess a double-barrelled structure: the expression that expresses the thought plus the expression that expresses the degree of assertion (or the colour of assertion). When it was attempted to turn these indicative sentences into questions it was only the expression of the thought that was acted upon; the expression of assertoric force or assertoric colour was left dangling. Hence it is just this double-barrelled structure that prevents certain expressions from being turned into questions.

However, it is very important to raise a contrast with Brian Loar's examples of buried assertoric force. Loar's examples contain double-barrelled assertions. Consider

(12) Jones who has seen the Vice Chancellor has resigned

together with the question

Has Jones who I know saw the V.C. really resigned?

The string 'who I know saw the V.C.' has *persistent* assertoric force; (12) consists of two assertions in the indicative mood. Now why is

Has Jones hopefully resigned?

rather awkward? Consider the indicative sentence 'Jones has hopefully resigned'. Let us spell 'hopefully' out in a separate clause.

Jones has resigned the chair. At least so I hope.

What must be understood is this : it is not just that you have a further bit of language expressing the speaker's hope; rather, you partly withdraw the original assertion; in asserting one you partly withdraw the other. So the *original* assertion cannot go into a question. This is very different from Loar's case. Examples (6)—(9) are double-barrelled in a different way; they contain both the expression of a thought and the expression of how this thought is asserted.

It appears then that many sentences contain conventional expressions which indicate that they are being asserted and how they are being asserted.

These ideas may possibly be applied to the problems of (3). There is a sort of feeling that in an *if . . . then* sentence the *then*-clause is more asserted than the *if*-clause. It appears to be a convention in the expression of assertion that the clause carrying the most assertion will be the one to which a modifier is usually attached. This may be part of the reason why (1)—(2) are acceptable and (3) sounds strange.

Either . . . or is symmetrical, both disjuncts being unasserted, and so refuses many of the modifiers we have examined.

The modifiers of assertoric force, however, do not behave exactly like those of assertoric colour.

Consider some *that*-clauses.

(13) Smith says that Jones is no doubt a maverick
(14) Smith says that Jones is regrettably turning out to be a maverick.

In (13) does 'no doubt' convey the attitude of the actual speaker or the reported speaker? It latches on, it seems, to the reported speech. On the other hand in (14), which deals with the colour of the assertion, we do not know for whom Jones' true character is regrettable. Smith could have made the report on Jones approvingly and the speaker of (14) could have supplied the 'regrettably' on his own account. Or, both the speaker of (14) and Smith could have regretted the way Jones was turning out.

Thus, we get a clue as to whose attitude to the degree of assertion is being conveyed; we do not so readily know who colours an assertion in reported speech. This might be a difference of convention between degree of assertion and colour of assertion.

It may be tentatively concluded then from this feeling of unease at such examples as

If no doubt he sets out early, he will avoid the rain

that speakers of a language have a practical grasp of the conventions of making an assertion. It is quite likely, however, that the conventions governing the expression of assertoric force have important differences from those governing assertoric colour.[1]

Sunderland Polytechnic

MATERIAL IMPLICATION, THE SUFFICIENCY CONDITION, AND CONDITIONAL PROOF

By Peter Gibbins

IN ANALYSIS 34.3 A. J. Dale defends the view that material implication constitutes an adequate paraphrase of 'if, then'. His defence, if it were successful, would be both interesting and elegant. For he argues that this view amounts to no more than the view that the schema of exportation is a logical truth. Exportation, he claims, can be shown to be a logical truth by examining its substitution instances. He concludes that we must accept that material implication is an adequate paraphrase of 'if, then', despite its being truth-functional and paradoxical.

Dale's defence is as follows.

First, the *sufficiency condition* for the falsity of 'if p then q' (namely that 'if p then q' is false if 'p' is true and 'q' is false) is sufficient to demonstrate the logical truth of the following eleven axiom schemata.

1. $(A.B) \rightarrow A$
2. $(A.B) \rightarrow B$
3. $[(A \rightarrow B) . (A \rightarrow C) . A] \rightarrow (B.C)$
4. $[(A \rightarrow B) . (B \rightarrow C) . A] \rightarrow C$
5. $[(A \rightarrow B) . A] \rightarrow B$
6. $A \rightarrow (A \lor B)$
7. $B \rightarrow (A \lor B)$
8. $[(A \rightarrow B) . (C \rightarrow B) . (A \lor C)] \rightarrow B$
9. $\sim \sim A \rightarrow A$
10. $A \rightarrow \sim \sim A$
11. $[(A \rightarrow B) . \sim B] \rightarrow \sim A$

The symbol '$\rightarrow$' stands for 'if, then', without prejudging the claim that material implication offers an adequate paraphrase of 'if, then'.

Secondly, the addition to these schemata of the schema of exportation

$$[(A.B) \rightarrow C] \rightarrow [A \rightarrow (B \rightarrow C)]$$

then yields, with modus ponens as a rule of inference, a complete set of axioms for the propositional calculus. Modus ponens is itself justified by the sufficiency condition. The defence of material implication therefore reduces to the defence of exportation, since the sufficiency condition is uncontroversial.

This elegant defence is unsuccessful.

First, though the sufficiency condition may be used to demonstrate the validity of rules of inference, it cannot be used, as Dale uses it, to demonstrate the logical truth of an axiom. Secondly, the addition of

exportation to the *rules of inference* justified by the sufficiency condition does not yield a complete set of rules for the propositional calculus. Thirdly, though Dale's defence fails, material implication may be defended along Dale's lines by adding the rule of conditional proof to those rules justified by the sufficiency condition, rather than by adding exportation. Fourthly, the rule of conditional proof is as controversial as material implication itself, since together with the rules justified by the sufficiency condition it yields those 'counterfactual fallacies' described by David Lewis (*Counterfactuals*, pp. 31–36). A Dale-type defence of material implication is successful if exportation is replaced by conditional proof, but conditional proof is no less controversial than material implication.

My first point is that though the sufficiency condition may be used to demonstrate the validity of a rule of inference, it may not be used to demonstrate the logical truth of an axiom. Take Dale's own example.

$$\text{If } p \text{ then } q,$$
$$\text{If } q \text{ then } r,$$
$$p$$

therefore $r.$

The sufficiency condition can be used to demonstrate the validity of the rule. For suppose r is false. Then the sufficiency condition shows that q is false whenever 'if q then r' is true. Similarly, if both conditionals are true, p is false. So if the conclusion is false, the assumptions cannot all be true, and the argument is therefore valid.

But this is not to say that the axiom schema

$$[(A \to B) . (B \to C) . A] \to C$$

is logically true, i.e. true for all substitutions of truth values of A, B and C. For the sufficiency condition tells us nothing about those cases in which either A or B is false. In fact, the sufficiency condition is compatible with '$p \to q$' always taking the truth-value 'false', in which case the axioms 1 to 11 are all contradictions.

However to each of Dale's axioms there corresponds a valid rule of inference. For example, axiom 8 corresponds to

$$\text{If } p \text{ then } q,$$
$$\text{If } r \text{ then } q,$$
$$\text{Either } p \text{ or } r,$$

therefore $q.$

And each of these rules is justified by the sufficiency condition. It would be better therefore to consider Dale's problem in the context of natural deduction and to ask: does the addition of exportation to the set of rules justified by the sufficiency condition yield a set which is complete for the propositional calculus?

My second point is that for Dale's eleven rules the addition of exportation does not yield a complete set, as may be seen from the following argument (which is essentially Copi's proof of the incompleteness of his nineteen rules in *Symbolic Logic* pp. 47-50).

Construct the following truth-tables for the truth-values o, 1, 2. '|p|' stands for the 'the truth-value of "p" '.

$$|\sim p| = 2 - |p|$$
$$|p.q| = \max(|p|, |q|)$$
$$|p \vee q| = \min(|p|, |q|)$$
$$|p \rightarrow q| = \min(2 - |p|, |q|)$$
$$|p \leftrightarrow q| = \max(|p \rightarrow q|, |q \rightarrow p|).$$

Dale's eleven rules are all hereditary with respect to taking the value o. That is, whenever all the assumptions take the value o, the conclusion takes the value o. Exportation (in both directions) is a valid rule of replacement, since its antecedent and consequent always take the same value for each assignment of truth-values to its component propositional variables. And yet there are classically valid arguments not hereditary with respect to taking the value o, i.e. not derivable from the eleven rules plus exportation as a rule of replacement. For example (Copi p. 50)

$$A \supset B$$
therefore $\quad A \supset (A.B).$

For when $|A| = 1$ and $|B| = 0$, $|A \rightarrow B| = 0$ but $|A \rightarrow (A.B)| = 1$. I conclude that Dale has not succeeded in showing that his rules, plus exportation, yield a complete set of rules for the propositional calculus.

However, another question arises. What can we add to the rules justified by the sufficiency condition in order to arrive at a complete set of rules for the propositional calculus?

My third point is that adding the rule of conditional proof is sufficient. For then one can derive Dale's axioms from the associated rules, and one can derive exportation from the rules of modus ponens and conditional proof alone. Dale shows in his paper how to demonstrate the completeness of these axiom schemata with modus ponens. A Dale-type defence of material implication reduces to a defence of the rule of conditional proof. This is an interesting result. Yet we should ask: does conditional proof require a defence?

My fourth and final point is that it does. A Dale-type defence of material implication must cover counterfactual conditionals. But applications of conditional proof, along with some other rules all of which are justified by the sufficiency condition, yield the counterfactual fallacies listed by Lewis. The fallacy of *strengthening the antecedent* follows from Simplification, modus ponens and conditional proof. The *fallacy of transitivity* (or hypothetical syllogism) follows from modus ponens

and conditional proof. The *fallacy of contraposition* follows from modus tollens and conditional proof.

Conditional proof is as much in need of defence as is material implication.[1]

University of Delaware

PLANTINGA AND LEIBNIZ'S LAPSE

By ROBERT BURCH

A theist might well be expected to embrace the claim, usually attributed to Leibniz, that God could have created any possible world in which he exists. For first intuitions suggest that this claim is implied by God's omnipotence. Alvin Plantinga, however, has derided the claim by calling it 'Leibniz's Lapse'.[1] Plantinga's motives are theistically sound: his view is part of a strategy for bolstering the free will defence against the problem of evil. However, in virtue of the affiliation between 'Leibniz's Lapse' and the doctrine of God's omnipotence, even if the affiliation is not a logical implication, it is advisable for the theist to exercise extreme circumspection before rejecting Leibnizianism as an error. I would like here to defend Leibnizianism by showing that Plantinga's argument against it fails. The reason it fails is simply that it begs the question: it assumes the existence of a certain possible world whose existence the Leibnizian would deny. In the process of showing this, I shall not question Plantinga's interesting interpretation of God's ability to create particular worlds. Rather I shall draw out several implications of his interpretation and use these in support of Leibnizianism.

I

Plantinga's argument against Leibnizianism is supposed to show that if there is free will, then there are possible worlds which contain God's existence, but which God cannot create. The primitive notion in the argument is that of God's (strongly) actualizing a state of affairs. When God actualizes a state of affairs, what he does is directly to bring it about that the state of affairs be actual. Where S is a state of affairs, Plantinga represents by 'GS' the state of affairs consisting of God's

actualizing S. He seems to accept as axiomatic about GS that a state of affairs S contains a state of affairs S' if and only if GS contains GS'.

Plantinga's basic definition, that of the notion of God's being able to create a possible world W, appeals to the counterfactual conditional (which, after Plantinga, I represent by using a single arrow). Thus, this notion is interpreted to mean that there is a state of affairs C such that GC is possible and $GC \rightarrow W$. I shall also say in this event that God can create a possible world W *by actualizing* C. It is upon this counterfactual interpretation of God's ability to create a world that Plantinga's argument against 'Leibniz's Lapse' chiefly depends. The argument uses indifferently either Stalnaker or Lewis semantics for the counterfactual conditional: it turns out that the logic of the case is such that the differences between Stalnaker and Lewis semantics are irrelevant. Stalnaker analyzes $\ulcorner A \rightarrow B \urcorner$ as: either A is impossible or else in the nearest (most similar) possible world (to the actual one) in which A obtains, B obtains too.[2] Lewis analyzes $\ulcorner A \rightarrow B \urcorner$ as: either A is impossible or else there is a world W in which both A and B obtain and such that in any world as near as W (to the actual one) in which A obtains, B obtains too.[3]

The core of Plantinga's argument depends on there being a possible state of affairs in which someone named Curley is offered a bribe and is free to accept or reject it. Let C be this state of affairs. And let A be the state of affairs consisting of Curley's accepting the bribe, $\bar{A}$ his not accepting it. Then Plantinga assumes that one of the following is true:

(16) $GC \rightarrow A$

(17) $GC \rightarrow \bar{A}$.

He argues that if (16) is true, then there is a possible world W^*, containing C and $\bar{A}$, which God cannot create. But if (17) is true, then there is a possible world W, containing C and A, which God cannot create. Since either (16) or (17) is true, there must be at least one possible world God cannot create.

The part of this argument that I want to focus on is the transition from the truth of (16) to the impossibility of God's creating W^*. For the transition from the truth of (17) to the uncreatability of W is exactly parallel and needs no separate treatment. I have no quarrel with the assumption that either (16) or (17) is true. Let us, therefore, look at the argument Plantinga provides that is supposed to show that if (16) is true, then there is a world W^* that God cannot create. This argument falls into two main sections. The first goes this way:

If (16) is true, 'then on the Stalnaker and Lewis semantics there is a possible world W such that GC and A hold in W, and such that A holds

[2] Robert Stalnaker, 'A Theory of Conditionals,' in Nicholas Rescher, *Studies in Logical Theory (American Philosophical Quarterly*, supplementary monograph, 1968), p. 98.
[3] David Lewis, *Counterfactuals* (Oxford: Basil Blackwell, 1973), ch. 1, Section 1.3.

in any world as close where GC holds' (p. 181). Then, let T be the largest state of affairs that God actualizes in W. It follows that

(18) $GT \rightarrow A$.

However, since A involves a free action, GT does not include A. 'So there is another possible world W^* where God actualizes the very same state of affairs as he does in W, and in which Curley rejects the bribe. W^* therefore includes GT and $\bar{A}$. That is, in W^* God strongly actualizes T but no state of affairs properly including T; and in W^* $\bar{A}$ holds.' (p. 181). In the second section of the argument Plantinga attempts to show that God could not have actualized this world W^*.

Before looking at the second section, let us pause to notice that the argument is already on dubious ground. Notice the sentence in which Plantinga says, 'That is, in W^* God strongly actualizes T *but no state of affairs properly including T*' (my italics). Where has Plantinga shown this? His use of 'That is' suggests that he thinks that this follows from the two preceding sentences. But it does not. The two preceding sentences argue that there is another possible world W^* (*another* in the sense that $W^* \neq W$) in which God actualizes T and in which $\bar{A}$ obtains. Very well. But the fact that in W^* God actualizes T does not mean that in W^* God actualizes *no more than* T. And the fact that T is the largest state of affairs God actualizes *in* W does not mean that T is the largest state of affairs he actualizes *in* W^*. Perhaps in W^* God actualizes not only T but something larger, i.e. a state of affairs S which includes T without being included in T. Plantinga provides no reason to think that this could not be so. And, as we shall see, a Leibnizian would argue that this must be so.

Let us now proceed to the second section of the argument. Suppose that God could have created W^*. Then there is a state of affairs C^* such that GC^* is possible and

(19) $GC^* \rightarrow W^*$.

Since W^* includes GT, we have

(20) $GC^* \rightarrow GT$.

Clearly, W^* includes GC^*. But now, $GC^* \,\&\, GT = GT$. Hence, from (18),

(21) $GC^* \,\&\, GT \rightarrow A$.

But from (20) and (21) it follows, on both Stalnaker and Lewis semantics, that

(22) $GC^* \rightarrow A$.

But, since A precludes W^*,

(23) $GC^* \rightarrow \overline{W^*}$.

Now, (19) and (23) can both be true only if GC^* is impossible. Thus the assumption that God could have created W^* leads to a contradiction and must be abandoned.

But notice that the crucial step in this section of the argument is the claim that GC^* & $GT = GT$. Plantinga's argument for this claim is that 'T, furthermore, is the largest state of affairs God actualizes in W^*; T, therefore, includes C^* and GT includes GC^*' (p. 182). Thus, the crucial claim in the second section of the argument depends precisely on the proposition that T is the largest state of affairs God actualizes in W^*. This, as we recently saw, is a proposition Plantinga provides no grounds for believing; and, as we shall shortly see, it is a proposition the Leibnizian would quickly deny.

II

The claim that T is the largest state of affairs God actualizes in W^* is a crucial, although unsupported, point of Plantinga's argument against 'Leibniz's Lapse'. I would like now to examine what a Leibnizian would say about this claim. In doing so I would like to draw out more fully Plantinga's notion of God's ability to create a world. I shall do this by proving four results about this notion.

To begin, let us take note of the fact that if a possible state of affairs S includes a possible world W, then S is identical with W. For a possible world is a maximal state of affairs.

With this in mind, let us work out the meaning of God's being able to create a possible world W, first on Stalnaker semantics and then on Lewis semantics. Let us suppose that there is a possible state of affairs C such that GC is possible and $GC \rightarrow W$. On Stalnaker semantics this means that in the nearest possible world (to the actual one), call it W', which contains GC, W is also contained. Since W' contains W, $W' = W$. This means simply that W itself is the nearest possible world (to the actual one) that contains GC. On Lewis semantics 'GC is possible and $GC \rightarrow W$' means that there is a world, call it W', containing GC and W, which is such that any world as close as W' (to the actual one) and containing GC also contains W. Once again, since W' includes W, $W' = W$. Moreover, every world containing W is identical with W. So, on Lewis semantics 'GC is possible and $GC \rightarrow W$' means that W contains GC and that any world as near as W (to the actual one) and containing GC *is* W, hence that W is the unique nearest possible world (to the actual one) containing GC. Thus, the meaning of God's ability to create a world W is the same for both Stalnaker and Lewis semantics: namely that there is a possible state of affairs C such that W is the closest possible world (to the actual one) containing GC.

Having elicited this much, we are now in a position to prove four results about God's ability to create a world, interpreted as Plantinga

does. The first result is this. Let W be a possible world that God can create, and let T be the largest state of affairs that God actualizes in W. Then God can create W by actualizing T. For, suppose that there is some state of affairs C such that GC is possible and $GC \rightarrow W$. Then W is the nearest possible world (to the actual one) containing GC. But now GT is possible since it is contained in the possible world W. Moreover, since T contains C, GT contains GC, and thus any world containing GT contains GC. It follows that W is the nearest possible world (to the actual one) containing GT; for, if there were a nearer world W' containing GT, then *per absurdum* there would be a nearer world containing GC. Hence $GT \rightarrow W$.

The second result is this. Let W be a possible world that God can create by actualizing a state of affairs C. Then God *cannot* create any other possible world W' (i.e. with $W' \neq W$) by actualizing C (and nothing larger than C). That is to say, suppose GC is possible and that we have both $GC \rightarrow W$ and $GC \rightarrow W'$. Then $W = W'$. For, since $GC \rightarrow W$, W is the closest possible world (to the actual one) containing GC. But, since $GC \rightarrow W'$, W' is the closest possible world (to the actual one) containing GC. This situation is possible only if $W = W'$.

The third result is this. Let W be a possible world that God can create. Let T be the largest state of affairs that God actualizes in W. And let W', with $W' \neq W$, be any other possible world containing GT. Then, if God can create W', he must in W' actualize a state of affairs T' that is strictly larger than T (i.e. T' contains T but T does not contain T'). For, from the first result we know that $GT \rightarrow W$. And from the second result we know that it is not the case that $GT \rightarrow W'$. Now, let T' be the largest state of affairs that God actualizes in W'. Since by hypothesis God can create W', we know from the first result that $GT' \rightarrow W'$. Now, since T' is the largest state of affairs that God actualizes in W', and since in W' God does actualize T (since W' contains GT), it is clear that T' contains T. But, if T contained T', then T would be identical with T', and thus GT would be identical with GT'. This cannot be the case, however, since if it were, then (because $GT' \rightarrow W'$) we would have that $GT \rightarrow W'$, which we already know not to be the case. Thus T does not contain T'. Hence, T' is strictly larger than T.

The fourth result is similar to the third. It is this. Let W and W' be possible worlds that God can create. And let T and T' be the largest states of affairs that God actualizes in W and W', respectively. Then $T = T'$ if and only if $W = W'$. For, taking up the 'only-if' condition, suppose that $T = T'$. By the first result above, $GT \rightarrow W$ and $GT' \rightarrow W'$. But since $T = T'$, $GT \rightarrow W'$. Then, by the second result above, $W = W'$. The 'if' condition follows directly from the definition of T and T'.

The third and fourth results show what, from the Leibnizian's point of view, is wrong with Plantinga's argument that God cannot

create every possible world in which he exists and in particular W^*. The argument begs the question by assuming a proposition that the Leibnizian should rush to deny: namely, that the largest state of affairs God actualizes in W^* is T, namely the very same state of affairs that is the largest state of affairs God actualizes in W. The Leibnizian should say that if W^* is a world which is different from W and in which God actualizes (at least as much as)T, then it must be false that T is the largest state of affairs that God actualizes in W^*. For the Leibnizian would maintain that God can create both W and W^*. It follows by the fourth result that the largest state of affairs God actualizes in W must be different from the largest state of affairs he actualizes in W^*. And by the third result it follows that the largest state of affairs he actualizes in W^* must be strictly larger than the largest state of affairs he actualizes in W. Another way of putting the Leibnizian's point is to say that W^*, *as Plantinga describes it,* is not a possible world at all. There is no such possible world because, on Leibnizian assumptions, no possible world world could satisfy all the characterizations Plantinga makes of W^*. One might, for merely rhetorical purposes, say that God could not create W^* as Plantinga describes it. But one could say this, not because there is a possible world W^* that God cannot create, but rather only because there is no such possible world as Plantinga's W^* at all.

Texas A & M University   © ROBERT BURCH 1979

LEWY'S CONJECTURES ABOUT TAUTOLOGICAL ENTAILMENT

By MICHAEL CLARK

I

IN his recent book *Meaning and Modality* Casimir Lewy conjectures that a certain logical relation which he calls 'strict entailment' is equivalent to Anderson and Belnap's tautological entailment. Lewy first introduces a notion he calls 'truth-functional equivalence', which I shall symbolize thus: ' $\triangle$ '.

> Df.: $A \triangle B$ iff $A \equiv B$ is a substitution instance (SI) of some tautology $\mathbf{A} \equiv \mathbf{B}$ with no tautologous proper parts.

The relation is manifestly non-transitive. Thus we have $p \triangle pq \vee p\bar{q}$ and $pq \vee p\bar{q} \triangle p$ $(q \vee \bar{q})$, since the corresponding material equivalences are respective SIs of $p \equiv pq \vee p\bar{q}$ and $pq \vee pr \equiv p(q \vee r)$, both of them tautologies with no tautologous proper parts. On the other hand we clearly do not have: $p \triangle p(q \vee \bar{q})$. Since therefore it is not an equivalence relation, and certainly not ordinary truth-functional equivalence, I prefer to use a less misleading name, and propose to call it ' $\triangle$-matching'. (Lewy himself called it ' $\triangle$-equivalence' in an earlier paper (1958).)

Lewy offers the following definition of his notion of strict entailment (symbolized here as ' $\rightarrow$ '):

> $A \rightarrow B$ iff $A \supset B$ is an SI of some tautology $\mathbf{A} \supset \mathbf{B}$ such that
> (1) $\mathbf{B}$ is purely contingent,
> (2) $\mathbf{\bar{A}}$ is purely contingent.
> (Lewy (1976), p. 150; notation adapted.)

But his definition is vitiated by being framed in terms of 'pure contingency', for according to Lewy a purely contingent wff, A, is one which is not only contingent (i.e. neither tautologous nor inconsistent[1]), but does not $\triangle$-match any conjunction with a (main) tautologous conjunct. As Lewy has agreed in correspondence, this notion of pure contingency proves to be vacuous, and with it strict entailment as defined above. There are no purely contingent wffs, and therefore no strict entailments. For any contingent wff A, $A \triangle$-matches $A(A \vee \bar{A})$, since $A \equiv A(A \vee B)$ is a tautology with no tautologous proper parts.

Lewy conjectured (1) that strict entailment was transitive and (2) that it was equivalent to tautological entailment (p. 150). Since there are no strict entailments, the second conjecture is false, and the first, though technically true, is only trivially so.

---

[1] Except in the final footnote, *inconsistent* means *truth-functionally inconsistent*, where 'truth-functionally' has its normal sense.

II

What Lewy was attempting to do in defining strict entailment was to restrict the non-transitive relation of 'S-necessitation' (adapted from Smiley (1958–9)) in order to make it transitive. S-necessitation (so-named by Lewy, who symbolizes it by '$\xrightarrow{s}$') holds between A and B iff

$A \supset B$ is an SI of some tautology $\mathbf{A} \supset \mathbf{B}$ with contingent $\mathbf{A,B}$.

Thus, for example, $p\bar{p}\xrightarrow{s}p(\bar{p} \vee q)$ and $p(\bar{p} \vee q)\xrightarrow{s}q$, but not $p\bar{p}\xrightarrow{s}q$. Again, we have $p\xrightarrow{s}p \vee \bar{p}$, and even $p\xrightarrow{s}p(q \vee \bar{q})$, but not $p\xrightarrow{s}q \vee \bar{q}$.

Now if we revert to Lewy's definition of strict entailment, it looks as if his clause (1) should require, not that $\mathbf{B}$ should fail to $\triangle$-match *any* wff with a tautologous conjunct, but that it should fail to $\triangle$-match any wff which has a tautologous conjunct *not S-necessitated by* $\mathbf{A}$. This would disqualify $p{\rightarrow}q \vee \bar{q}$ and $p{\rightarrow}p(q \vee \bar{q})$, whose corresponding tautological entailment wffs are false, while retaining $p{\rightarrow}p(p \vee \bar{p})$, for example, which corresponds to a true tautological entailment wff. Indeed, we might try putting the point solely in terms of S-necessitation: $\mathbf{B}$ must not S-necessitate any tautology like $p \vee \bar{p}$, $q \vee \bar{q} \vee r$, etc. unless it is also S-necessitated by $\mathbf{A}$. Analogous considerations apply to clause (2). A third clause is needed to secure transitivity.

Accordingly I offer the following definition of a reconstructed notion of strict entailment (retaining the symbol '$\rightarrow$' for it).

$A{\rightarrow}B$ iff $A \supset B$ is an SI of some tautology $\mathbf{A} \supset \mathbf{B}$ with contingent $\mathbf{A,B}$, such that, for every *primitive* tautology T and every *primitive* inconsistent wff I,

(1) if $\mathbf{B}\xrightarrow{s}T$, then $\mathbf{A}\xrightarrow{s}T$;
(2) if $I\xrightarrow{s}\mathbf{A}$, then $I\xrightarrow{s}\mathbf{B}$; and
(3) if $I\xrightarrow{s}\mathbf{A}$ and $\mathbf{B}\xrightarrow{s}T$, then $I\xrightarrow{s}T$.

A *primitive* tautology is a tautologous disjunction of atoms (propositional variables or their negations), a *primitive* inconsistent wff an inconsistent conjunction of atoms. Lewy has told me that he believes this redefinition succeeds in capturing the notion he was after.

I have a proof, not given here, that the relation just defined is equivalent to tautological entailment,[1] and consequently that it is transitive. Recast in terms of this revised notion of strict entailment, both of Lewy's conjectures are demonstrably true.

[1] Tautological entailment, symbolized here by '$\xrightarrow{t}$', holds only between truth functions. By definition: $A\xrightarrow{t}B$ iff there is a true wff $A_1 \vee . . .\vee Am\xrightarrow{t}B_1$ &. . .& Bn, where the antecedent is a disjunctive normal form of A, the consequent a conjunctive normal form of B, both normal forms being obtainable as described on pp. 156–7 of Anderson and Belnap (1975); and the latter is true iff, for every i, j, A_i shares an atom with B_j. The transitivity of the relation is proved on p. 160. Wffs like $r\bar{r} \vee pq\xrightarrow{t}(q \vee r)(p \vee \bar{p})$ are false: hence the need for clause (3) of the redefinition above, which prevents corresponding wffs from holding as strict entailments.

I have included the qualification 'primitive' at two places in the re-definition since, although the relation defined without it can be shown to be sufficient for tautological entailment, I cannot prove that it is necessary.

In Chapter 9 of his book Lewy argues, rightly in my view, that no reconstruction of our preformal notion of entailment which rejects the principles of the disjunctive syllogism (MTP, etc.) is acceptable. If this is right, then both T-entailment and redefined strict entailment are damned as candidates for that role, for it is well known that disjunctive syllogism does not hold for the former. Its failure for the revised notion of strict entailment can, in fact, be established without difficulty directly from the definition.

The original notion of strict entailment was introduced in a more general form in Lewy (1958), along with a weaker notion which is defined for the case of truth functions on p. 151 of his book and called 'strict necessitation' ($\underset{\to}{n}$). Since this relation is also defined in terms of purely contingent wffs it is as vacuous as the original strict entailment, but may be reconstructed analogously by taking the redefinition of strict entailment and omitting clauses (2) and (3). The reader may check that, on this revised definition, $p \vee q\bar{q} \underset{\to}{n} p$ but not $\bar{p} \underset{\to}{n} \overline{p \vee q\bar{q}}$, so that the relation is non-contrapositive—as Lewy believed it was on his definition. As he would agree, this makes it scarcely more plausible as a reconstruction of our preformal notion of entailment.

III

Lewy points out that S-necessitation results from putting a sieve on strict implication to exclude those implications which hold solely in virtue of the modal status of antecedent and consequent. And it is clear from the fact that redefined strict entailment is equivalent to T-entailment that the latter, the Anderson–Belnap notion, in its turn puts a sieve on S-necessitation. It is not even necessary to establish the equivalence to show this, since Anderson and Belnap have shown (pp. 219–220 (1975)) that $A \underset{\to}{t} B$ only if $A \underset{\to}{s} B$.

S-necessitations include cases like (i) $p\bar{p} \vee q \underset{\to}{s} q$, (ii) $q \underset{\to}{s} q(r \vee \bar{r})$ and (iii) $p\bar{p} \vee q \underset{\to}{s} q(r \vee \bar{r})$, i.e. implications containing an intrusive, irrelevant inconsistent disjunct in the antecedent, or an irrelevant tautologous conjunct in the consequent, or both. T-entailment traps cases like these in its sieve. Indeed it traps all those cases in which a conjunctive normal form of the consequent contains an irrelevant tautologous conjunct and/or a disjunctive normal form of the antecedent contains an irrelevant inconsistent disjunct. Thus $A(\bar{A} \vee B) \underset{\to}{s} B$ is caught in the sieve: $A\bar{A} \vee AB$, a dnf of the antecedent, contains the intrusive disjunct $A\bar{A}$. Or, put in terms of the redefinition of strict entailment: any tautologous

implication with contingent antecedent and consequent which has $A(\bar{A} \vee B) \supset B$ as an SI will have the form $\mathbf{A}(\bar{\mathbf{A}} \vee \mathbf{B}) \supset \mathbf{B}$; but whereas $\mathbf{A}\bar{\mathbf{A}} \xrightarrow{\cdot} \mathbf{A}(\bar{\mathbf{A}} \vee \mathbf{B})$ we do not have $\mathbf{A}\bar{\mathbf{A}} \xrightarrow{\cdot} \mathbf{B}$.

IV

After many years of trying to capture our 'preformal notion of entailment' Lewy has concluded that it is inconsistent, and that it is less counterintuitive to identify it with strict implication than (for example) to reject the disjunctive syllogism (like Anderson and Belnap) or to reject transitivity (like Geach).

As I have indicated, I agree that Anderson and Belnap's way out is too preposterous to countenance. But Lewy's conclusion, however reluctant, fails to deal with a puzzling problem. It is vividly illustrated by Anderson and Belnap (p. 332). The first edition of Quine's *Mathematical Logic* contained a system in which the Burali-Forti paradox was found to arise, although Cantor's paradox was apparently avoided. But if a contradiction entails any wff whatsoever, how can we construe the claim that Cantor's paradox was avoided? Moreover, we do not allow in mathematics and logic that a given theorem ordinarily has *any* theorem as its consequence: we think of some theorems as yielding certain others, but we do not think that every theorem yields every other. Any answer to the problem of entailment should either endorse and explicate this intuitive assumption or satisfy us that it is misguided.

It would indeed be quite absurd to deny that derivation or deduction in logic and mathematics was transitive. This is why Geach has distinguished *entailment* from *derivation* when claiming that entailment is non-transitive. A derivation is sound only if each step is a valid entailment-step, but entailment need not in general hold, he thinks, between the premisses and the final conclusion. Until recently, Geach has (in effect) identified propositional calculus entailment with S-necessitation.

If we regard S-necessitation as too weak a relation for entailment, on the grounds that we do not wish to admit entailments like (i)—(iii) in section III with their intrusive, irrelevant non-contingent components, then Geach's recently suggested revision in ANALYSIS 35.6 (1974-5, p. 187, q.v.) seems little better, for it admits the entailment of $r \vee p(q \vee \bar{q})$ by p, among indefinitely many like cases.

Now the notion of $\triangle$-matching suggests a way of defining a non-transitive relation of entailment which excludes all cases involving irrelevant non-contingencies while admitting the principles of the disjunctive syllogism. I shall use the symbol '$\xrightarrow{\delta}$' for the new relation, of which I offer the following definition:

$A \xrightarrow{\delta} B$ iff $A \supset B$ is an SI of some tautology with no non-contingent wf proper parts.

The relation is obviously a decidable one (cf. Geach (1972), pp. 187–8).

In the present context I shall not attempt to defend its identification with entailment against possible objections I am aware of, nor shall I discuss it any further here except to make two brief points.

Firstly, the relation lies between S-necessitation and tautological entailment: it is obvious that $A \overset{\delta}{\to} B$ only if $A \underset{\to}{s} B$, and I have been able to prove that $A \underset{\to}{t} B$ only if $A \overset{\delta}{\to} B$.

Secondly, if the relation is appropriately generalized to apply to quantificational logic,[1] then it should hold between a conjunction of (finitely many) axioms of Quine's original ML system and the contradiction involved in the Burali-Forti paradox, but not between any such conjunction and the contradiction in Cantor's paradox.[2]

[1] We may generalize the relation by letting $A \overset{\delta}{\to} B$ hold iff the closure of $A \supset B$ is an SI of a valid closure of some wff $A \supset B$ a closure of no proper part of which is valid or inconsistent (in the sense of *logically false*). Cf. George Myro's generalization of S-necessitation, ANALYSIS 32.2 (1971–2), p. 55.

[2] I am indebted to Dr. Alan Rose of the Nottingham Mathematics Department for stimulating discussions which helped me to discover the results of section II.

University of Nottingham   © MICHAEL CLARK 1979

REFERENCES

A.R. Anderson and N.D. Belnap, 'Tautological Entailments', *Philosophical Studies* 13 (1962).
——*Entailment, Vol. I* (1975).
P.T. Geach, *Logic Matters* (1972).
——'On Entailment', ANALYSIS 35.6 (1974–5).
C. Lewy, 'Entailment', *PASS* 32 (1958).
——*Meaning and Modality* (1976).
George Myro, 'A Note on Strict Implication and Entailment', ANALYSIS 32.2 (1971–2).
T. J. Smiley, 'Entailment and Deducibility', *Proceedings of the Aristotelian Society* 59 (1958–9).

'EVEN IF'

By Allen Hazen and Michael Slote

ONE of the minor vexations of the Vexed Question of the Conditional is the question of 'even if' conditionals. All kinds of conditionals—indicative, would-counterfactual, and might-counterfactual—have 'even if' variants,[1] and logicians have told a wide variety of stories about them. Some have claimed that 'if' and 'even if' conditionals differ in meaning. Thus the Anderson–Belnap school of relevant logicians, notoriously, reject the (classically and intuitionistically valid) schema

$$(P \rightarrow (Q \rightarrow P)),$$

calling it "Positive Paradox." When it is urged against them that instances of this schema, such as 'if justice is to be done, then, even if the sky falls, justice is to be done', seem intuitively to be perfectly valid, they sometimes reply[2] 'But that has an "even if", so it isn't really an instance of Positive Paradox.' Others—David Lewis in *Counterfactuals* comes to mind—have held that 'if' and 'even if' conditionals of the same kind have the *same* truth conditions, and that the function of 'even' is to comment on some *pragmatic* feature of the context.[3] This latter approach makes a stronger and bolder empirical claim about the language, and we will follow it here.

'Even if' conditionals are but one instance of a more general use of 'even'. Roughly, 'even' is inserted before a constituent of a sentence (which may itself be embedded in a longer sentence) to mark that the sentence, though thought to be true, is felt to be more unexpected or less evident than some other sentence, differing from it only in the marked constituent, suggested by the context. Thus, for example, the 'even' in 'Even Joe Camp went to hear Feigl's talk' suggests that Camp is (antecedently) less likely than, say, Larry Laudan to attend philosophy of science lectures. 'Even' can come before the subject noun phrase, as here, or before the verb phrase, or before an adjective in predicate position ('He was red in the face. He was even purple'). Often, but by no means always, the other sentence than which the 'even'-marked one is less likely-seeming is suggested by the context in a very straightforward way: by occurring in the same stretch of discourse. When it does, the two sentences may be conjoined. When they are, the usual

[1] In *Fact, Fiction and Forecast* (Indianapolis: Bobbs-Merrill, 2nd edition, 1965, ch. 1), Nelson Goodman offers an account of 'even if' subjunctive conditionals that doesn't seem to apply to 'even if' might-counterfactuals. Such an account has a strike against it if a competing account can treat all 'even if' conditionals in a unified manner.

[2] In conversation. The idea that 'even' affects the truth conditions of conditionals can also be found in Goodman *op. cit.*, and in John Pollock, 'Four kinds of conditionals', *American Philosophical Quarterly* 12, 1975, pp. 51–9.

[3] See *Counterfactuals*, Cambridge, Mass: Harvard University Press, 1973, p. 33n.

rules about deletion of identical constituents in conjoint structures apply: 'After touching Jesus' robe the cripple walked and even ran'. Or the marked sentence and its less unexpected congener may be the disjuncts of a disjunction, with similar deletions: 'The letter might not come until Friday or even Saturday.'

Interpreting the 'even if' conditionals in terms of this general function of 'even' seems, for the usual Ockhamite reasons, preferable to making up a special story about them alone.[1] It seems to work. If you say 'If you drank a gallon of whiskey, you'd get drunk', I can reply 'Even if I (only) drank a quart of whiskey, I'd get drunk'. The first conditional is more evident, on the principle that if whiskey makes you drunk, drinking more whiskey is surer to make you drunk, and, though the truth value of the 'quart' conditional surely doesn't depend on it, the 'even' is appropriate only because the context invites comparison with the "surer thing" 'gallon' conditional.

David Lewis's suggestion,[2] that 'even' marks the failure of the consequent to contrast with the truth or with the consequent of some other conditional, is consistent with ours but less informative. Failure to contrast is a symmetric relation, so he has no explanation why, if you started off with 'If you drank a quart of whiskey, you'd get drunk', it would be inappropriate for me to reply 'Even if I drank a gallon, I'd get drunk'—although I *can* say 'If I drank a gallon, I would (also) (certainly) get drunk.'

The constituent marked with 'even' in an 'even if' conditional is the adverbial phrase consisting of 'if' followed by the antecedent. There is independent evidence that the 'even' must be thought of as attaching to this phrase rather than to the whole conditional: conditionals can also be written with the 'if'-phrase following the consequent—'I'd come to the party if Jenny were going to be there'—and when they are the 'even' still gets inserted right before the 'if': 'I'd come to the party even if she weren't.'

It is the unexpectedness of the whole conditional that is being marked, though, and not the unexpectedness or "farfetchedness" of the antecedent. In the whiskey example, the antecedent of the marked version is less far-fetched than that of the unmarked, but in the following example it is the other way round: 'If Jones had been fired, he wouldn't have bought the boat. Even if he'd been fired and found a new job the next day, he wouldn't have bought the boat'. Similar examples can be found with might-counterfactuals.

---

[1] The 'even's in the examples above, like those in 'even if' conditionals ,would be translated 'auch' in German and 'même' in French. Since 'even', 'auch', and 'même' have very different meanings elsewhere, this is evidence that the uses typified by these examples form a natural family with the use in 'even if' conditionals, and that a common explanation is in order.

[2] See *Counterfactuals*, p. 33n.

Since the characters in our whiskey example could easily be abstemious, it should be clear that 'even' can be used in true counterfactuals with false consequents. However, in many cases, 'even' is inserted in a counterfactual when the consequent is true. This is explained by assuming that (when it *is* true) one of the alternative "counterfactuals" most obviously true and most salient in a given context is the one with the antecedent 'things are the way they actually are'. There is more to be said for this bit of sophistry than at first meets the eye: not only does it explain why 'even' is so often appropriate in counterfactuals with true consequents, it can also explain why in some cases it isn't. Suppose a certain sort of elementary particle always does X in circumstance A, but that in circumstance B it is genuinely indeterminate whether one will do X or Y. Suppose that in a certain experiment condition B holds, but that the particle does X anyway. Then the counterfactual 'Had circumstance A held, the particle would (certainly) have done X' is true, but the corresponding 'even if' counterfactual is inappropriate. The reason is that in this case the 'Had A obtained' counterfactual is evidently true, being supported by a law of quantum physics, but the corresponding true-antecedent conditional, 'Had condition B held, as it in fact did, the particle would have done X,' is, given the indeterminacy of the outcome under condition B, far from evident.[1]

Two last points. First, an open question: What exactly has to be the same in the 'even'-marked and unmarked sentences? In the whiskey example, 'you' changes to 'I', keeping the reference the same. In other examples the grammatical form of the consequents of the two conditionals is the same, but the reference is different: 'If the steamer trunk had been on the bed, I would have noticed it. Even if (only) the briefcase had been on the bed, I would have noticed it.' Similar problems arise with the other uses of 'even' described in the second paragraph.

Finally, a grammatical point that we think *must* be relevant to something. At least with indicative conditionals, the form

If P, and if Q, then R

is ambiguous. It can be read either as

(If P then R) and (if Q then R)

or as

---

[1] In his "official" system of counterfactual logic, David Lewis requires that any counterfactual with true antecedent and consequent be true. This is consistent with saying that some of them, such as the one in our example, are unobvious or implausible. He also considers an alternative system, based on allowing more than one world in the innermost sphere of similarity around a world, on which some such counterfactuals may be false. Examples such as ours, in which the antecedent and consequent are both true but the truth of the antecedent fails to make it determinate that the consequent will be true, provide what is perhaps the best reason for accepting this latter logic.

If (P and Q) then R/If P, then (if Q then R).[1]

'Even' disambiguates it. The form

If P, and even if Q, then R

can be read only as

(If P then R) and (Even if Q then R).

[1] These two paraphrases seem to represent a single reading, and to be paraphrases of each other. Note that '((P & Q)→R)' and '(P→(Q→R))' are equivalent in classical two-valued logic and in intuitionistic logic, but not in David Lewis's counterfactual logic or the Anderson–Belnap relevant logics.

Trinity College Dublin   © ALLEN HAZEN *and* MICHAEL SLOTE 1979

CARTESIAN SCEPTICISM AND EPISTEMIC LOGIC

By MARK STEINER

IN one of the most celebrated philosophical passages, the first of his six *Meditations*, Descartes claims to have demonstrated a sceptical conclusion concerning the senses by arguing that the very sensations which seem to confirm the existence of an "external" world might be occasioned by a dream:

> I am a man, and . . . consequently I am in the habit of sleeping, and in my dreams representing to myself the same things or sometimes even less probable things, than do those who are insane in their waking moments. How often has it happened to me that in the night I dream that I found myself in this particular place, that I was dressed and seated near the fire, whilst in reality I was lying undressed in bed! At this moment it does indeed seem to me that it is with eyes awake that I am looking at this paper; that this head which I move is not asleep, that it is deliberately and of set purpose that I extend my hand and perceive it; what happens in sleep does not appear so clear nor so distinct as does all this. But in thinking over this I remind myself that on many occasions I have in sleep been deceived by similar illusions, and in dwelling carefully on this reflection I see so manifestly that there are no certain indications by which we may clearly distinguish wakefulness from sleep that I am lost in astonishment. . . .
> Now let us assume that we are asleep and that all these particulars, e.g. that we open our eyes, shake our head, extend our hands, and so on, are but false delusions; and let us reflect that possibly neither our hands nor our whole body are such as they appear to us to be.[1]

Descartes goes on to argue that his argument leaves intact geometrical knowledge, which is possible even in one's dreams; but he evidently

[1] Descartes, *Meditations on the First Philosophy*, in *The Philosophical Works of Descartes*, translated by Elizabeth S. Haldane and G. R. T. Ross (New York: Dover, 1955), pp. 145–146.

thinks that the argument enables him to conclude that he does not know that he is seated by the fire; in symbols,

(1) $-K(S)$.

To reach this conclusion, Descartes asserts the following two premises:

(2) $D \rightarrow -K(S)$

('If I'm dreaming, then I don't know I'm sitting'), and

(3) $-K(-D)$

('I don't know I'm not dreaming').

Proposition (2) is the weakest construal of Descartes' claim that, if we sleep, bodily particulars are 'false delusions' and that 'possibly neither our hands nor our whole body are such as they appear to us to be'. The basis for this claim is the very nature of dreaming—even if Descartes happens to be dreaming that he is lying in bed, he is not credited with knowledge of that fact. (3) is another matter; it is evidently supposed to follow from Descartes' remark that 'manifestly' there are 'no certain indications' by which we may distinguish the waking from the sleeeping states, given his rule that nothing is to be asserted as known unless it *is* certainly indicated.

Let us take (2) as self-evident and (3) as definitely proved. The purpose of what follows is to enquire what *other* premises, if any, are needed to prove (1). If Descartes had an argument for

(4) D ('I am dreaming')

the conclusion would follow by modus ponens from (2). But Descartes is not claiming to dream, but that he cannot be sure he isn't—Proposition (3). Contraposing (2), however, we get

(5) $K(S) \rightarrow -D$,

and by the well-known necessitation rule of epistemic logic

(6) $K(K(S) \rightarrow -D)$;

if one has proved (5), one has thereby proved that one knows it. By another rule of epistemic logic, we conclude from this

(7) $KK(S) \rightarrow K(-D)$.

Now suppose that we postulate for Descartes the 'KK-principle' of Hintikka[1], to wit:

(8) $K(S) \rightarrow KK(S)$.

Then, from (7) and (8),

(9) $K(S) \rightarrow K(-D)$.

[1] Jaako Hintikka, *Knowledge and Belief* (Ithaca: Cornell University Press, 1962).

Contraposing again,

$$(10) \quad -K(-D) \rightarrow -K(S).$$

Remembering that we accept (arguendo) Descartes' claim to have proved (3), we can *now* detach (1) from (10) by modus ponens. QED.

Thus, besides Propositions (2) and (3), the argument of *Meditation I* seemingly rests on the *KK*-principle. Informally: I don't *know* whether or not I'm dreaming, so I don't know for sure that I *know* I'm sitting here—for if I *knew* that I know this, I'd know I'm not dreaming. Since I don't know that I know (that I'm sitting), then I don't know it either, by Hintikka's principle.

If our reconstruction of Descartes' argument is sound, it can easily be answered by denying the *KK*-principle. Indeed, there may be other reasons to deny it. As David Shatz points out, on a Causal Theory of Knowledge the *KK*-principle licenses the dubious inference from

The fact that $K(S)$ causes the belief that $K(S)$

to

The fact that $KK(S)$ causes the belief that $KK(S)$.[1]

As Shatz points out, the Causal Theory of Knowldge (and other such theories) presuppose a conception of the knowing subject as a mere reliable reflection of his environment. On such theories, even to require that the subject believe what he knows may seem superfluous, for belief is not the only reflection of one's environment. Shatz cites the example of dogs, often credited with knowledge, but seldom with beliefs; and, the example of students, credited with knowledge of test answers they have little confidence in.

Since theories like the Causal Theory dispense with the *KK*-principle, may they claim to refute Cartesian scepticism? Not necessarily; another epistemic principle recommends itself, weaker than the *KK*-principle, but strong enough to sustain scepticism. The principle is suggested by the necessitation rule used above, and reads as follows:

(*)   If one is committed to $-K(P)$ ['*P*' for any sentence], then it is irrational of him to assert *P*.[2]

This principle resembles, but is weaker than, the claim that statements like

It's raining but I don't know it

are "self-stultifying," "contradictory," or the like.

Returning to (7), that is

$$KK(S) \rightarrow K(-D),$$

by contraposition we get

(11) $-K(-D) \rightarrow -KK(S)$.

Once again relying on (3), we detach

(12) $-KK(S)$.

Without the *KK*-principle, Descartes can no longer prove $-K(S)$. But principle (*) makes it irrational, nevertheless, to assert $K(S)$ ('I *know* S to be the case'). For Descartes is committed to $-KK(S)$ on the basis of a claimed proof, and so long as he continues to do so, he cannot rationally assert $K(S)$. That is, from the fact that (12) is *proved* (not merely from its truth), it follows that

(13) Asserting $K(S)$ is irrational.

This conclusion should satisfy the sceptic. But we can go further with the stronger epistemic principle

(**) If it is irrational to assert $K(P)$, then it is irrational to assert P.

Though this principle is indeed stronger, it too has intuitive plausibility. For the assertion of S is tantamount to the assertion of $K(S)$ (the assertion of S is properly followed by 'How do you know?' and the answer 'I don't' appears to be a withdrawal of the original bald assertion). In any event, if (**) is accepted, then from (13) we conclude

(14) The assertion of S is irrational.

And this is quite a sceptical conclusion indeed. It is, however, based upon a concept of rationality which might be questioned as presupposing a social context and thus the falsity of scepticism. Short of this questioning, the only way to fight scepticism, so it seems to me, is to attack (3) directly. Dropping the *KK*-principle is not enough.[1]

Columbia University

ALSTON AND SELF-WARRANT

By Douglas Odegard

WILLIAM ALSTON ('Self-Warrant: a Neglected Form of Privileged Access', *American Philosophical Quarterly*, Oct. 1976) claims that beliefs about the intrinsic nature of one's own current conscious states are self-warranted in the sense that they are justified just by virtue of being beliefs of that sort. Such beliefs (called *B*'s) satisfy the latter condition only if (i) they are reliable as such, and (ii) no specific sorts of *B*'s are significantly more reliable than *B*'s in general. Condition (ii) is added to avoid the need to specify something further about a given *B* in order to justify holding it at a suitable level of justification.

Alston claims that *B*'s are self-warranted on the ground that they are rarely falsified, even though independent tests are possible. Sometimes he holds that they are never falsified, although he says that he needs only the weaker premise in order to establish his conclusion. The sufficiency of the weaker premise is doubtful, however. If *B*'s are occasionally falsified, we can describe some of the tests which reveal the errors. Any *B* which passes such tests might then be significantly more reliable than *B*'s in general, in which case (ii) is violated. For example, suppose that *S* reports feeling pain, that a physiologist claims that there is nothing physically wrong with *S*, and that a psychiatrist says that *S* is sufficiently neurotic to mistake non-painful sensations for painful ones. If in the circumstances this evidence is good enough to falsify *S*'s belief, we might say that *B*'s which are not opposed by psycho-physiological evidence are significantly more reliable than *B*'s in general. If so, then someone justifiably holds a *B* only if there is no such evidence, and *B*'s are not *self*-warranted.

Of course, whether we should say this depends upon how rare such occasions are and what 'significant' means. If the frequency of error is in the order of one in a million and significance is understood statistically, the difference is not significant. But if this is what significance amounts to, the "privilege" of our access to our own conscious states is no different from the way anyone stands to matters about which he is rarely mistaken; *e.g.* from the way trained chicken sexers stand to the sex of chicks, or from the way gifted mathematicians stand to elementary calculations. Indeed, if someone is sufficiently careful in forming beliefs about the conscious states of others, his success rate might be high enough to give him an access to others which is just as "privileged" as his access to his own states.

Alston may be trying to avoid this consequence when he argues: 'I would suggest that when we employ these independent checks, we rarely find them at variance with the subject's report; where they are,

the rarity of this will justify us in either opting for the self-report, or putting its falsity down to insincerity rather than mistake. Thus insofar as we have independent empirical evidence, we have no reason to think that B's are ever mistaken' (p. 269). This could be taken to imply that, although independent checks occasionally produce negative results, the very rareness of such occasions allows us to dismiss the results as worthless. In that case Alston is not conceding that B's are occasionally falsified, but only that occasionally they confront negative evidence which turns out to be worthless. But such evidence does not lose its worth if the circumstances explain how the given subject can be holding a mistaken B in spite of the fact that the vast majority of B's are correct. In the above example of the neurotic, the psychiatric evidence explains how the subject can believe that he feels pain even if he does not. It thereby explains how he can go wrong in holding a B even though most B's are correct. This allows the physiological evidence to acquire value and the latter can then combine with the remaining evidence to falsify the subject's belief (*e.g.* it might combine with a lack of non-linguistic pain behaviour and with a subsequent memory that there really was no pain). It is doubtful that Alston does want to argue in favour of saying that B's are never falsified. But if he does, his argument is ineffective.

Suppose, however, that he does not and that he is prepared to accept the consequences of allowing occasional falsifications of B's. Are the falsifications rare enough to be insignificant, *i.e.* to establish no significant difference between the reliability of B's in general and the reliability of the specific B's which manage to pass independent tests? Alston argues that they are by claiming that the only 'remotely plausible' mark of a falsified B is really just a mark of a failure to believe. If he is right, there is no way to characterize falsified B's as opposed to unfalsified ones and hence no way to specify a subspecies of B's which would be more reliable than B's in general. The mark he mentions is a subject's not possessing all his faculties, or his being 'too confused, distracted, or whatever, to "know what is going on" ' (p. 270).

The notions of faculty-privation and confusion *are* vague enough to cover cases of failures of belief. But they can also cover cases of mistaken beliefs. For example, someone with a poor grasp of geometry might see both a square and a circle, correctly believe that their perimeters are the same, and mistakenly believe that they have the same areas. Someone in a state of philosophical confusion might believe that what he sees is a sense datum, in a sense in which there are no sense data. Mistakes of this sort are sufficiently concerned with the "intrinsic" nature of conscious states to count as B's. They can be characterized. And they occur frequently enough to be significant.

Similarly, although normally we are sufficiently inattentive that we do not form beliefs about our current conscious states, distraction can

DISHONEST RELATIVISM

By B. C. Postow

IN this paper I shall argue that a sort of moral position which is fairly common among philosophers and other thoughtful people is in some sense dishonest. The sort of moral position with which I am concerned has two components: (1) some view according to which people sometimes morally ought to do certain things, and (2) some meta-ethical view according to which a normative view that conflicts with one's own well grounded normative view may itself be a well grounded or acceptable moral view. The first component is held by almost all people with moral convictions[1]; the second component is held by a smaller but still considerable number of people who, despite their moral convictions, are impressed with the seeming impossibility of finding an objective foundation for one correct moral view. Many philosophers concur in this; the grounds on which they take their own normative views to be rationally justified would justify conflicting normative views as well as their own. John Rawls, to take a well known example, considers a normative theory to be justified if it represents a reflective equilibrium between our moral intuitions and the demands of theory.[2] But, as Rawls himself points out, conflicting normative theories might represent reflective equilibrium for persons with differing moral intuitions. 'Even should everyone attain wide reflective equilibrium, many contrary moral conceptions may still be held'.[3] The fact that this situation would deprive him of any grounds for saying that a theory contrary to his own is mistaken does not disturb Rawls, for 'the procedure of reflective equilibrium does not assume that there is one correct moral conception'.[4]

The question which I would like to consider here is whether it is legitimate to continue to accept one's *own* normative view in such a situation. I am willing to grant that it is legitimate to continue to believe that one's own normative view is well grounded in such a situation, for a meta-ethical theory like Rawls' defines 'well-groundedness' in such a way that conflicting normative views may in fact both be well grounded. But accepting a normative view involves more than believing that it is well grounded. (If it did not, one would have to accept conflicting views which one believed to be well grounded.) Accepting a moral view involves, in addition to the belief that the view is well

---

[1] It is possible to have moral convictions about what ought to be the case without having any moral convictions that anyone ought to do anything. See I. L. Humberstone, 'Two sorts of "Ought"'s', ANALYSIS 32.1 (1971–2), pp. 8–11.

[2] *A Theory of Justice* Cambridge, Mass: Harvard University Press, 1971, pp. 48–50.

[3] 'The Independence of Moral Theory', *Proceedings and Addresses of the American Philosophical Association*, 48 (1974–75), 9.

[4] Ibid.

grounded, some sort of practical commitment to the view. It is the legitimacy of this practical commitment which I wish to question in cases where one believes that a rival normative view is equally as well grounded as the view which one accepts.

It is notoriously difficult to characterize the practical commitment inherent in accepting a normative theory, but we can say enough to make the problem clear. Included in a whole-hearted acceptance of a normative view is at least some sort of minimal assent to certain prescriptions which I shall call the prescriptions that correspond directly to the view. If P is a normative view which is itself taken to be a prescription, then the prescription which 'corresponds' directly to it will simply be P itself. If P is the judgment (or judgment analogue) that lying in certain circumstances is always wrong, the directly corresponding prescription would prohibit lying in these circumstances. If P is the judgment that Fred ought not to lie, the directly corresponding prescription forbids Fred to lie. Now what is involved in 'minimal assent' to a prescription? It seems reasonable to say that A gives at least minimal assent to a prescription which commands something of B (who may or may not be identical with A) only if A would be willing to use the prescription or its equivalent to offer guidance to B in appropriate circumstances, assuming that A is not especially timid. By 'appropriate circumstances' I mean circumstances in which, in A's opinion, it would be proper and prudent to offer sincere guidance to B. For example, if Alice assents to a prescription directing Bob to return the money which he has embezzled, then she should be willing in appropriate circumstances to advise Bob (or exhort him, or whatever) to return the money. If she is not willing to do this in appropriate circumstances, this indicates that she does not assent to the prescription. My thesis requires that the acceptance of a normative view involves at least this sort of minimal assent to the prescriptions which correspond directly to the view.

It may be thought that where P is a judgement about a particular person's obligations, someone who accepts P must assent not only to the directly corresponding prescription, but also to the universal prescription which corresponds indirectly to P. For if I think that a particular person ought to do a particular act in certain circumstances, then consistency requires that I also think that all people in circumstances which are defined as relevantly similar by my moral view ought to do acts which are defined as relevantly similar by my moral view. But although consistency may require this, it seems that I could be confused and inconsistent in not holding the generalized view, yet still be said to accept the particular judgment. In this case, I would assent to the prescription directly corresponding to P, but I would not assent to the prescriptions that are [quasi] implied by P. I would not object

to the specification of a stronger sense of 'accept' in which a person
would be said to accept a set of judgement/prescriptions P only if that
person also accepts all judgement/prescriptions that are [quasi] implied
by the members of P, perhaps together with descriptions of nonmoral
facts; but the problem with which I am concerned can be stated in
terms of the weaker sense of 'accept'.

Let me illustrate why it seems illegitimate for a person to accept
a normative theory while believing that a conflicting theory is equally
well grounded. Suppose that Alice's moral theory implies that Bob
(and other people in ordinary circumstances) ought to abstain from
animal products. Because she assents to the corresponding prescription,
Alice is willing in appropriate circumstances to use this prescription to
advise or exhort Bob to abstain from animal products. But Alice
believes that Bob's moral theory, which condones and even commands
the use of animal products, is equally well grounded as her own.
Therefore she believes that Bob has no moral reason to abstain from
animal products. She may also believe that he has no nonmoral reason
to abstain, in that abstinence would not serve any of his nonmoral goals.
By hypothesis, Alice would be willing in such a case to advise or exhort
Bob to do something which, in her opinion, he has no reason to do—
and this seems in some sense unreasonable or dishonest.

Now Alice might not agree that Bob has no moral reason to abstain
from animal products. She might point out that although *his* moral
theory does not provide a reason for him or anyone to abstain, *her*
moral theory provides a reason for everyone to abstain. And surely it
would be correct to cite the fact that Bob has an obligation to do some-
thing as a reason for him to do it when one is advising or exhorting him
to do it. Thus if Alice's moral theory is well grounded, Alice can cite
Bob's obligation according to this theory as a reason for him to abstain
from animal products.

There is an easy reply to Alice's objection in the case where Bob's
moral theory implies an obligation to use animal products. For if his
obligation according to Alice's well grounded theory to abstain from
animal products constitutes a reason for him to abstain, then his obliga-
tion according to his own well grounded theory to use animal products
constitutes an equally good reason for him not to abstain. These reasons
cancel each other out. Thus Bob has no all-things-considered reason to
abstain from animal products. Since Alice should realize this, she would
be unreasonable or dishonest to exhort Bob to abstain in such a case.

Let us now consider the case where Bob is merely permitted rather
than obligated to use animal products according to his own moral
theory. It would by hypothesis be equally reasonable for him to accept
his own view and reject Alice's, as it would be for him to accept Alice's
and reject his own. Now I assume that it would not be even more

reasonable to accept both his own view and Alice's conflicting view rather than to accept one and reject the other. Thus rejecting Alice's view would be an optimally reasonable position. And if he takes this position, he can reasonably refuse to accord any weight at all to moral reasons provided by Alice's view. But if it would be equally as reasonable for Bob to refuse to accord any weight at all to the only reason which Alice can cite for his abstaining from animal products as it would be for him to accord weight to this reason, it seems that the overall force of reasons cannot be said to favour Bob's abstaining from animal products over his using them. The guidance which Alice is prepared to offer, however, explicitly favours Bob's abstaining over his using animal products. So once again the guidance which she is prepared to offer is not supported by the overall force of reasons.

Of course Alice may never actually be in a situation in which she is willing to offer someone guidance which fails to be supported by the overall force of reasons; but she must be *prepared* to do so in appropriately specified hypothetical situations. And so must the holder of any normative view which ascribes obligations to agents, if she or he also holds that a contrary normative view may be equally well grounded (and if that fact would not make her or him give up the original normative view). Now in order to specify what the holder of any particular normative theory, such as Rawls', must be prepared in hypothetical situations to exhort people to do, it would be necessary to delve into the intricacies of the particular normative theory in question. But it seems enough reason to call a position dishonest, that a person who adopts it must be prepared in *some* hypothetical situation to advise or exhort someone to do *something* which, according to the position, there is no all-things-considered reason (including reasons furnished by morality) for that person to do.[1]

[1] Thanks to Gilbert Harman, Thomas Nagel and the Editor of ANALYSIS for their helpful comments and suggestions.

HOW NOT TO MAKE A NEWCOMB CHOICE

By André Gallois

ON being confronted with two boxes, labelled Box One and Box Two, and being allowed to choose between taking Boxes One and Two and taking Box One alone, I am supplied with the following information, which I know to be correct: no matter what I choose, there is $1,000 in Box Two, but a predictor has predicted the choice I will make and has placed $1,000,000 in Box One if and only if it predicted that I would take Box One alone. Moreover, I have strong inductive grounds for thinking that the predictor will turn out to have made a correct prediction. In this situation, what is it rational for me to choose?

This problem, devised by William Newcomb, has recently been subjected to a great deal of discussion in the literature. George Schlesinger (e.g. 'The Unpredictability of Free Choices', *British Journal for the Philosophy of Science*, 1974) attempts to show that, assuming that there could be a predictor of human choices, decisive considerations can be brought in favour of choosing Box One alone, and equally decisive considerations can be brought in favour of choosing Box One and Box Two. Don Locke, in 'How to Make a Newcomb Choice' (ANALYSIS 38.1, January, 1978) appears to accept Schlesinger's argument for choosing to take both boxes, but rejects the argument for choosing to take Box One alone. In this paper, I will try to show three things: first, that Locke gives no good reason for thinking that the possibility of reverse causation is irrelevant to the problem posed by Newcomb; second that the argument for taking both boxes, called the well-wisher argument, is misconceived; and finally, that Locke's reason for rejecting the argument for taking Box One alone either embodies a familiar modal fallacy or else presupposes that compatibilist accounts of free choice are incorrect.

I

Locke opens his discussion of the arguments for choosing both boxes by observing that 'The case for Choice Two [taking both boxes] seems straightforward enough: the Predictor has already laid his money down—or not, as the case may be—so I cannot be worse off if I take both boxes'. Locke goes on to remark that even if we can best explain the Predictor's success by postulating precognition (which Locke takes to involve reverse causation), this is irrelevant to the question of which choice is the most rational one. Locke's argument for the irrelevance of reverse causation to Newcomb's problem goes as follows: even if reverse causation is possible, it is not possible to alter the past, so even if reverse causation is possible, it remains true that either the Predictor

has placed $1,000,000 in Box One or he has not. Hence, the argument for Choice Two is unaffected by the possibility of reverse causation.

At least two things may be meant by saying that a present choice alters a past prediction. The first is that the choice is a sufficient condition for a certain prediction having been made in the past. The second implies bringing it about that a certain event which did occur did not occur. If this is in the sense in which Locke thinks that altering a past prediction would amount to 'making it other than it was', then we may agree with him that the suggestion that past predictions are alterable is incoherent. However, ruling out the possibility of altering the past in this sense does not demonstrate that reverse causation is irrelevant to Newcomb's problem. To see why, let us consider a situation that, following precedents in the literature, we may call the Bmocwen situation.

The Bmocwen situation is like the situation in Newcomb's problem, except that an Observer will subsequently place $1,000,000 in Box One, if and only if the Chooser is observed to choose Box One alone. Now, it is clear that the causal interaction between the Chooser and Observer is relevant to the question of what is the rational choice in the Bmocwen situation. Taking his cue from Locke, though, the Bmocwen Chooser could deploy the following argument: I cannot alter the future. At best, as in the Newcomb situation, I can make a choice that is causally sufficient for the Observer placing $1,000,000 in Box One. However, the Observer will either place $1,000,000 in Box One or he will not, and in either case, I am better off taking both boxes.

Obviously this argument, though it parallels Locke's, fails to show the irrelevance of ordinary causal considerations to the Bmocwen problem. Whether or not reverse causation is a live option, Locke's argument is simply a standard argument for fatalism applied to the past instead of the future. However, Locke has an additional argument for the conclusion that, whatever the causal connection between the Predictor's predictions and the Chooser's choices, the Chooser ought to make Choice Two. This is Schlesinger's well-wisher argument to which I now turn.

Locke gives the following version of the well-wisher argument:

  · I can know that an intelligent, well-informed and perfect well-wisher, knowing what is in the boxes, would advise me to make Choice Two, both if Box One is full and if it is empty; what an intelligent, well-informed and perfect well-wisher would advise me to do must be in my best interests; so, whatever is in Box One, it must be in my best interests to make Choice Two.

It may be that, if I were in the Newcomb situation, a well-wisher would be unable to advise me or, if I am countersuggestible, would advise me to do the opposite of what he would want me to do. At any rate, whatever a well-wisher would advise me to do, the important question

is what a well-wisher would want me to do. With this point in mind, let us ask whether the well-wisher argument would apply with equal force in the Bmocwen situation.

Suppose that at t_1 the Observer either places or refrains from placing $1,000,000 in Box One, depending on the choice the Chooser was observed to make. At t_2 the Chooser is permitted to inspect and remove the contents of whichever box he can lay claim to in accordance with his earlier choice. Prior to making his choice, the Chooser reasons as follows: if a perfect well-wisher should become apprised of the contents of Box One and Box Two in the interval between t_1 and t_2 what will he want me to have chosen? If the well-wisher discovers that Box One is empty, he will want me to have chosen both boxes, for I will then end up with $1,000 instead of nothing. If he discovers that Box Two contains $1,000,000 he will want me to have chosen both boxes, for then I will end up with $1,001,000. So, no matter what Box One contains the well-wisher will want me to have chosen to take both boxes. Hence, it is irrational for me to take Box One, even though the Observer will promptly penalize me for failing to do so by refraining from putting anything in Box One.

If Locke were to reply that the well-wisher argument fails to apply to the Bmocwen situation precisely because a well-informed and intelligent well-wisher would realize that what is in Box One depends on the choice made at t_1, exactly the same can be said about the well-wisher in the Newcomb situation. I conclude that as the well-wisher argument is not a sound argument for making Choice Two in the Bmocwen situation, it is not a sound argument for making this choice in the Newcomb situation.

II

I now come to Locke's rebuttal of Schlesinger's argument for making Choice One. Locke lists the four outcomes relevant to the argument as

A Predictor predicts Choice One, Chooser makes Choice Two and gets $1,001,000;

B Predictor predicts Choice One, Chooser makes Choice One and and gets $1,000,000;

C Predictor predicts Choice Two, Chooser makes Choice Two and gets $1,000;

D Predictor predicts Choice Two. Chooser makes Choice One and gets nothing.

Given that the Predictor is infallible, of these four outcomes only B and C are possible. Therefore, it is rational for the Chooser to opt for Choice One if he has reason to believe in the Predictor's infallibility.

Of the two possible outcomes, B is preferable to C, and the Chooser's making Choice One is a necessary condition for the realization of B.

Locke notes that this argument for Choice One depends on taking the Predictor to be infallible. Still, even if the evidence clearly favours the Predictor's being infallible, Locke maintains that we have no argument for Choice One:

> . . . once the Predictor has made his prediction then either Outcome B or Outcome C will also be impossible, though the Chooser does not know which.

Consequently,

> . . . once the Predictor has made his prediction his infallibility, the impossibility of his going wrong, ensures that the predicted choice is the only choice available, and *a fortiori* the best choice available.

In what sense, if any, is either Outcome B or Outcome C impossible, given that the Predictor is infallible and has made a prediction? Granted that if Outcomes A and D are impossible and the set of outcomes A through D includes all possible outcomes, then we may infer both

> necessarily (if the Predictor predicts Choice One, then the Chooser makes Choice One and gets $1,000,000),

and also

> necessarily (if the Predictor predicts Choice Two, then the Chooser makes Choice Two and gets $1,000).

It would, however, be fallacious to argue from

> necessarily (if the Predictor predicts Choice One, then the Chooser makes Choice One and gets $1,000,000) and the Predictor predicts Choice One

to

> necessarily the Chooser makes Choice One and gets $1,000,000.

Arguing in this way would amount to committing the well-known modal fallacy: necessarily (if p, then q), p, so necessarily q.

No modal fallacy is involved if Locke is arguing that, since the Predictor's past prediction is fixed, the Chooser's choice is fixed as well. If this is Locke's argument, however, it does not show that once an infallible Predictor has, say, predicted that the Chooser will make Choice One, Outcome C is impossible in the same sense that Outcomes A and D are. Outcomes A and D could be discounted, because it is causally impossible for the Predictor to go wrong. However, to admit that if the Predictor has predicted Choice One, it is fixed that the Chooser will make Choice One, is compatible with denying that Outcome C is

causally impossible, since it is causally possible for the Predictor to have made a different prediction. Consequently, contrary to Locke, we do not have the same reason for claiming that either Outcome B or C is closed to the Chooser that we have for claiming that Outcomes A and D are closed. Hence, Locke has not given a convincing argument for the claim, '. . . given that the Predictor is absolutely infallible, it is at the time of choosing equally impossible, and in just the same sense [as it is impossible for Outcomes A and D to be realized], for the Chooser to make any other choice than that predicted'.

It may seem that it is nevertheless in some sense impossible for the Chooser to make any choice other than the one he makes. Be that as it may, the central issue is whether it follows from the fact that the Chooser's making a certain choice is fixed, that the Chooser is not free to make any other choice. It is a characteristic compatibilist thesis that the obtaining of causally sufficient conditions in the past for my making a certain choice in the present does not, by itself, restrict the range of choices open to me. Any compatibilist worth his salt would insist on distinguishing between the Predictor's compelling the Chooser to make Choice One and the Predictor's prediction simply being a causally sufficient condition for the Chooser's making Choice One. Without a supplementary argument for the falsity of compatibilism, Locke's discussion of the argument for Choice One fails to demonstrate its inadequacy.

Even if Locke produced a decisive refutation of compatibilist approaches to free choice, a case could still be made for Choice One. While it would be pointless for the Chooser to speculate about the consequences of his realizing Outcomes A and D, since he knows that he cannot realize either A or D, it would not be pointless for him to speculate about the consequences of realizing B and C. Though the Chooser knows that either he cannot make Choice One or he cannot make Choice Two, he does not know which of these choices is closed to him. Consequently, there is still a point to the Chooser's weighing the relative desirability of Outcomes B and C and making his choice accordingly.

In summary: Locke has not made a case for the irrelevance of reverse causation to the Newcomb problem. In addition, the central argument he gives for making Choice Two, the well-wisher argument, is unsound. Finally, the argument for Choice One is unimpaired by the considerations that Locke brings against it.

Monash University   © ANDRÉ GALLOIS 1979

WOODRUFF ON REVERSE DISCRIMINATION

By JOHN ROBERTSON

PAUL Woodruff has claimed that discrimination (understood as the according of special treatment on grounds of group membership, where this is irrelevant to the question of whether or not the treatment is deserved) is wrong only when it is part of a pattern creating a group whose respect is thereby diminished unfairly (ANALYSIS 36.3, 38.1.) This is an important and interesting thesis. If it is right, those disadvantaged by reverse discrimination cannot claim to have been wronged, for reverse discrimination does not unfairly reduce the respect of a group to which they belong. Further, it offers a determinate goal for reverse discrimination which it is within the means of government policies to achieve. Governments can, by reverse discrimination, compensate for the diminished respect suffered by groups as a result of discrimination; they cannot by these means exact restitution from individuals guilty of discrimination in the past on behalf of those individuals who have suffered from it. But, on Woodruff's account, it is only the former that is the object of reverse discrimination (38.1, p. 64). My main target in this note is Woodruff's claim that reverse discrimination is, in this context, significantly different from restitution. But before I turn to this issue, there is a preliminary point, important to Woodruff's position, which I wish to address.

Woodruff argues that isolated acts of discrimination are not wrong. A private banker, he contends, who prefers a nephew over another, perhaps better qualified candidate, discriminates against the latter, but does nothing wrong, though perhaps something foolish. He would do something wrong, however, if his action contributed to a pattern which excluded all non-members of his family from banking careers. The difficulty with this is that as the first case is described, the position in the bank is a benefit which it is the banker's right to confer as he pleases. A father, without acting wrongly, may favour a wastrel son over more deserving progeny in his will, for if, after all, it is his fortune, it is his to bequeath to whomever he wishes, and this right overrides the general right of his offspring to fair treatment. Precisely the same right is exercised by the banker, and that his act of discrimination is not wrong is, I think, wholly dependent on this point. There are limits to this right, to be sure: it does not extend to the right to collude with others to deny whole groups access to worth-while jobs (Woodruff's second case), nor perhaps to the right to conform to a general practice which contributes to this result. Still, Woodruff's example does not show that isolated acts of discrimination are not unfair, but only that one may

sometimes have the right to ignore considerations of fairness and please oneself.

It is hard, however, to find a clear counter-example to the claim that isolated acts of discrimination are not wrong. To discriminate is to confer or withhold some benefit on irrelevant grounds, and that benefit is either one's own or it is not. If it is one's own, I have argued, then one's right to dispense it as one wishes overrides others' claim to fair treatment. If the benefit is not one's own, if one is an agent of someone else or a public official conferring a public benefit, one has no such right, and may not discriminate. But perhaps this is not because of the right to fair treatment on the part of the candidates for the benefit, but rather because such discrimination violates an obligation to respect the interests of the public or of a private benefactor. And of course, whether public officials may be required in the pursuit of the public interest to discriminate in distributing a public benefit leads us to the question we are trying to get clear about: may the public, through government policies, advance its interests by discriminatory means? It seems unlikely, then, that arguments against reverse discrimination based on the wrongness of the unfair treatment of individuals will be conclusive. Discrimination is surely unfair, but the strongest conclusion one may draw from this is that the unfairness of reverse discrimination is something to be said against it.

I shall argue that reverse discrimination is unfair, and sufficiently so to warrant a government searching for alternative means of compensation (or drastically modifying its commitment to the fair treatment of individuals). It is unfair, not because it involves discrimination, but because it fails to meet certain straightforward conditions on fair compensation between groups. Before I turn to my argument for this, I want to stress that any form of compensation must involve resources that are scarce and valuable. One cannot compensate groups or individuals for unfairly diminished respect, income, or property with sand or scrap paper. If groups are to be compensated for unfairly diminished respect, the resources will be those that contribute to restoring this respect, such as preferential access to professional schools and, by means of this, to worthwhile work. The increased scarcity of these goods for those who bear the costs of compensation is, as Woodruff says (38.1, p. 64; 36.3, p. 159), an evil separate from discrimination, but it is *not* an evil separate from compensatory discrimination, for the discrimination could not be compensatory without this. The distribution of the burdens and benefits of reverse discrimination are, then, relevant to the fairness of that program as a form of compensation.

To get clear on how seriously unfair reverse discrimination is it is helpful to consider an analogue to Woodruff's account of it. Corporations, if disadvantaged by the unfair practices of their competitors, may sue for damages. As a result of the unfair practices the

disadvantaged corporation may have had to reduce benefits to some of
its employees, but this would not give them a claim to the damages the
corporation collects. The damages may go entirely to members of the
corporation who did not suffer, or suffered least, from the unfair com-
petition. The corporation guilty of unfair practices may, in order to pay
damages, have had to reduce benefits to some of its employees, but it
would not be under an obligation to reduce the benefits of just those
employees responsible for the unfair practice. It may distribute the losses
as it will, and those who suffer from them cannot claim that, as they were
not responsible for the unfair practice, they should not bear the loss.
I think the analogy to Woodruff's account of reverse discrimination as a
form of compensation is precise: compensating groups for unfairly
diminished respect *is* restitution, between groups regarded as individuals
analogous to corporations, with respect rather than capital the good
exchanged.

Why is there no obligation on the part of a corporation to see that
those who have suffered least as a result of unfair practices do not
benefit most from compensation, or to see that those who have not con-
tributed to the unfair practice do not suffer most for it? The answer lies,
I think, in the contractual nature of the relationship between employees
and the corporations employing them. The employees have a counter-
vailing right to organize to fight reduction in benefits and to force a
fairer distribution of the burden. They have the right, if this fails, to
quit and look for employment elsewhere. Were it impossible for
employees to organize to protect themselves, and impossible as well
to quit their positions and work elsewhere, fairness would require
that governments compensate for their weakness by ensuring that in
restitution between corporations the burdens and benefits were spread
around. Now neither of the countervailing rights enjoyed by the em-
ployees in this example are available to those disadvantaged by reverse
discrimination. There is nothing analogous to a private corporation
against which they can organize to force a fair distribution of the burdens,
nor can they withdraw from their sexual or racial group. Nor will it do
to suggest, as Woodruff might, that since the burden involved is simply
diminished respect for the group as a whole, its distribution is guaran-
teed, unlike the burden imposed on a corporation by a damage suit.
For the diminished respect is wholly consequent on limiting access to
such scarce goods as higher education and worthwhile work, and these
burdens are not distributed fairly. I would conclude from this that
governments may, in an effort to compensate unfairly disadvantaged
groups, employ discriminatory means, if they can ensure that the burdens
on the groups discriminated against are fairly distributed, or if the
members of these groups, like the corporation employees, have rights
sufficient to enforce such a distribution. Unless these conditions are met,

governments cannot meet the legitimate claims individuals have on them for fair treatment. Since reverse discrimination meets neither of these conditions, it seems it is an unacceptable means of compensation.

One may agree with Woodruff, then, that reverse discrimination is not to be condemned just because it is discrimination. Indeed, if groups can owe and be owed compensation, and if the goods in terms of which compensation is most appropriate are such things as opportunities for professional training, then preferential treatment for individuals on grounds of group membership cannot be avoided. But reverse discrimination does deserve to be condemned as a procedure for compensation undertaken by a government on which individuals have a claim to fair and equal treatment. It is true that, where compensation is due, it is due whether there are fair means of making it or not. It is not, then, a decisive objection to reverse discrimination that it cannot, as things now stand, be fairly implemented. Were there no alternative, governments might be justified in turning to such a policy as a last resort, burdening themselves with a new debt as they pay off an old one. There are, however, other programs available, such as tax supported programs designed to raise the general level of education and skills amongst a group that has suffered discrimination. These are compensatory programs, for they involve distinguishing, among all those who would benefit from them, the claims of just those whose need derives from a history of discrimination. It may be objected that this proposal would merely shift the burden for compensation from, e.g. young white males, who can bear it fairly easily, to the poor, who cannot: the unfairness is preserved, and only the victims changed. But if these programs are tax supported, the burden is not shifted. The burden is the tax, and it can be made to fall exactly on that portion of the public thought to owe compensation. It is true that such a program may involve temporarily ignoring some of the needs of those who are poor but do not qualify for it. But we should ordinarily think that debts owed to one party take priority over the needs of another as claims on our resources. We are not justified in giving a sum of money owed to one person to another because the other needs it as much or more. If the history of discrimination has indeed created a social debt to some groups, then it has created a claim to our resources that takes priority over the claim of those who came to be disadvantaged through bad luck or folly. The latter would bear the cost of compensatory programs only in the sense that, were there no debts requiring such programs, more resources would be available to meet their needs. This does not, however, violate their claim to fair and equal treatment.

Syracuse University

MORAL EVIL WITHOUT CONSEQUENCES?

By M. J. COUGHLAN

IN arguing for the irrelevance of the Free Will Defence, Steven E. Boër claims that we could have a world in which persons have evil intentions but perform no actions having 'morally objectionable consequences'.[1] He holds that God could allow us freedom of will while intervening on all occasions when a person attempts to inflict harm, so that moral evil would be possible but would never result in physical or natural evil. Freedom of the will, Boër says, is a licence to choose and try, not a warranty of success.

Whatever the merits of the Free Will Defence, there are some difficulties in conceiving a world such as the one proposed here. There is, for example, a problem of credibility: agents would be forming various intentions to do certain actions, and trying to implement those intentions, in spite of everyday experience which ought to teach one that intentions of this kind were *always* frustrated. It is not easy to believe that persons would attempt murder by shooting (Boër's example) when no attempt at murder had ever been successful.

This difficulty might seem relatively harmless. It could be said that, if persons were dissuaded from trying to do evil, so much the better. But, in Boër's world, persons would not be dissuaded in the way in which a potential bank robber drops his plans when he sees that the security measures are such that his chances of success are minimal. The kind of dissuasion, if it can be called that, which would be operating is the kind which induces us not to try walking on air. Boër's persons are not free to inflict evil in the same sense in which we are not free to walk on air, and to claim that the essential element in our freedom is not impaired by this, because we are still free to try to walk on air, is unconvincing.

But the problems go far deeper. Consider the concept 'trying'. There are special circumstances in which a person might try to walk on air, for example, children or the insane might do so. But could we say of anyone who has learnt from experience that it is not possible to walk on air, that he is 'trying to walk on air'? Surely it is a condition of trying to do x that one does not believe that x is an impossible action for one, so that a person who attempts murder cannot think that murder is impossible?[2] If, therefore, persons continue to make evil choices in Boër's world much as they do in this one, it must be the case that they do not learn from their failures in this particular area. This, presumably, would

---

[1] The Irrelevance of the Free Will Defence', ANALYSIS 38.2, 110–2.
[2] This observation, and its consequences for the argument, were suggested to me by R.A. Sharpe.

be an *ad hoc* condition, decreed by the almighty God, to ensure that we retained the impression that we could inflict evil. One must assume that this remarkable condition will not apply universally, for, if we did not learn from our failures in general, we would have not just a modified world but an entirely different one, yet Boër's world is '. . . exactly like our own except that evil machinations would never result in harm to any innocent party' (p. 112). Furthermore, an interesting corollary of the imposition of this special condition would be that it would prevent us from knowing that God was good in the sense that he did not allow us to harm the innocent, that is, we would still be wrestling with this side of the problem of evil. And, in the fulness of time, when we might hope to discover the strange ways of the Divinity, there would arise the final difficulty of describing as 'good' a God who had systematically deceived his creatures. Descartes, for one, would be very disillusioned!

Finally, in Boër's world, it is difficult to see how any evil action or event would be instantiated, and, therefore, how we might have any criteria for the use of the word 'evil', that is, it is not clear how we might acquire the *concept* of evil. Without a concept of evil it would be meaningless to speak of moral perfection or imperfection, yet Boër concedes the common opinion that having free will is incompatible with the impossibility of moral imperfection. The Free Will Defence shows, he says, '. . . that God (logically) cannot create a world in which persons are free but never make morally reprehensible choices or form morally objectionable intentions' (p. 110). (This may be an overstatement. Surely it is not necessary that morally reprehensible choices are made, but merely that they are possible. But this point does not affect the issue because, in either case, persons in Boër's world must be able, at least, to possess the concept of evil.) Let us consider the problem which besets the acquisition of a notion of evil in this world.

Firstly, persons will never be successful in carrying out evil intentions, assuming that they can have them. This, in itself, is not an insurmountable difficulty. Nobody has been successful at walking on air but we can make sense of the concept. We know what air is, we know what it is to walk on something; we know what would *count as* walking on air. In Boër's world it is tempting to think that we could build a concept of evil intention or evil action in a similar way, that is, by putting together our general concept of intention and a notion of evil derived from observation of natural evil; for example, we see a tree fall accidentally, killing a rabbit, so we plan to have a heavy object fall at the appropriate time on our chosen victim. But Boër proposes a world in which all evil consequences and events that can be prevented without compromising free will *are* prevented. Natural evil, he argues, is the real problem for the theodicist because it is not logically bound up with

free will (even if one regards natural evil as the consequence of Satan's intentions, it can be disposed of in the same way as the evil consequences of any person's intentions). Boër's world, therefore, would contain no natural evil, so that we would be deprived of the model on which we might have built, or identified, evil intentions. If persons, or any other living thing, were never injured, much less killed, it is difficult to see how we could give meaning to an expression such as 'attempted murder'. The situation that Boër describes, that is, where one person picks up a gun, aims it at another's head, and pulls the trigger, meets certain criteria because actions of this description usually *result,* or could reasonably be expected to result, in the other person's head being blown off. The difficulty is to see how we could have reasonable expectations of such consequences in Boër's world. If we had no basis for these expectations, there would be no justification, indeed *no sense,* in describing these actions as cases of attempted murder, or even simply as the expressions of a murder wish.

In general, it seems that in a world where there was no natural evil and where no attempt to inflict evil ever resulted in morally objectionable consequences, or could reasonably be expected to so result, there would be no concept of evil, and, *a fortiori,* no concepts of trying to do evil or intending evil. If Boër and the Free Will Defence theodicists are right, free will would not be possible in this world. Whether this consequence is accepted or not will depend on one's understanding of free will. It might be objected that one would still be free to try to effect consequences which would be in fact evil, without recognizing them as such, and, therefore, one can retain free will without a concept of evil. Although such choices might be open to one, they are certainly not moral choices and, in the present context at least, it is clear that free will is regarded primarily as the freedom to make moral choices.

St David's University College,   © MICHAEL J. COUGHLAN 1979
Lampeter

BEING UNFREE TO AND BEING UNFREE

By George E. Panichas

IN a recent and valuable essay concerned with some of the complex issues pertaining to the relationship between liberty and the law, J. P. Day has offered and defended some revised and revitalized versions of the views of Hobbes and Bentham.[1] Professor Day argues carefully for an analysis of unfreedom, the implications of which stand to refute the contentions that the unfreedom of an agent to do or perform some act is logically dependent upon the agent's being incapable (cf. section 4, pp. 260–1), unreasonable (cf. subsection 5.1, p. 261), immoral (cf. subsection 5.3, pp. 262–3), or not feeling free (cf. subsection 5.4, p. 262) in, or with respect to, performing the relevant act.

While Day's arguments on these scores are both interesting and important on their own account (I shall not call any of them into question here), their purpose is to establish a foundation for what I believe is Day's central claim; namely, that a person is unfree to do or perform some act or other just in case that person is rendered retrievably unable to do so by the actions of another. And it is with respect to this conclusion that Day believes his analysis has significance with respect to the concept of liberty. Day concludes that '. . . the sufficient and necessary conditions of the truth of "A is unfree to D" is the truth of "B makes A retrievably unable to D by Eing A"' (p. 264–5)[2] and emphasizes his conclusion with the claim that '. . . whenever any assertion is made about liberty, it is always possible to discover exactly what (if anything) is being asserted by seeking values for the four variables in the above formula, in order to obtain a clear statement which is either true or false.' (p. 265.)

In what follows, I shall not argue that Day's analysis of 'A is unfree to D' is wrong; but I do want to call into question the belief that in analysing 'A is unfree to D' one has in fact given an analysis exhaustive of A's being unfree to D in the sense which is germane to A's liberty. My point will be that Day neglects the fact that while it may be true that on some occasions 'unfree to' is used synonymously with (or as an instantiation of) 'unfree', none the less, on other occasions, the two are quite distinct. And, in neglecting this fact, the acceptability of Day's remarks concerning assertions about liberty is severely limited.

---

[1] J. P. Day, 'Threats, Offers, Law, Opinion and Liberty', *American Philosophical Quarterly*. Volume 14, Number 4, October 1977, pp. 257–71. All page and section references appearing in parentheses refer to this article.

[2] The word 'retrievably' is a crucial here, Day argues (p. 264), because if a person is rendered unable to perform a certain act *permanently*, then a necessary condition of his being unfree to perform that act—that he is able to do so—cannot be met. Thus a person who has been made irretrievably unable to perform an act is not unfree to perform that act. In such cases, it is the case that the question of one's being free to do or perform some act cannot (properly) arise.

The basis of my objection can be seen in cases where we do not say that a person is unfree, or that his liberty has been reduced, but where that person has been made retrievably unable to perform some act because of the activities of another. Consider a fairly standard case of medical treatment: Mr. Adams, who suffers from chronic respiratory congestion, enlists the services of Dr. Brown who is, in time, successful in the treatment of Adams. At time T_1, the time prior to the completion of treatment, Adams was able to cough with sufficient vehemence so as to clear his lungs and afford himself relief. Also, at T_1, Adams could resist such coughing and continue to suffer. At time T_2, the time subsequent to the successful treatment of Adams by Brown, Adams is no longer able either to clear his lungs with vehement coughing, or to resist such coughing. At T_2, then, Adams is clearly unable to perform at least two acts which he was able to perform at T_1—and this is due to Brown's treatment of Adams. But Adams' condition is chronic, so that he may well (indeed probably will) suffer from this condition again. Brown's treatment of Adams, then, has made Adams retrievably unable to cough; or, as Day would have it, it is true that 'B makes A retrievably unable to D by Eing A'. Yet surely, while we might agree that Adams is unfree *to* cough so as to gain relief, or unfree *to* resist coughing, we do not want to say that Adams' liberty or freedom has been reduced or thwarted; indeed, we are inclined to say quite the opposite. Adams is unfree *to* cough or resist coughing at T_2, but he is neither unfree nor less free than he was at T_1. Adams' liberty is, I believe, quite intact.

It should be noted at this point that I am *not* claiming that Brown has not made Adams unfree *to* cough or resist coughing. In fact, I would agree with Day that neither Adams' desires or wishes (to be treated), nor Adams' reasons (or lack thereof) count as necessary conditions of the truth of 'Adams is unfree to cough' (cf. pp. 258–9, and 261). However, the question to be asked here is: what might one appeal to, in the above case, so as to justify the claim that Adams is both unfree to (cough) *and* unfree, if not to considerations of Adams' desires at T_1 and/or the reasonableness (or even, the desirability) of his situation at T_2? The problem with Day's account is that such factors are barred from consideration.

It is interesting to note that Day's analysis is confined to the sentence 'A is unfree to D' (from which he makes inferences about A's liberty), and that he does not consider for careful analysis the sentence 'A is free to D' (or even, 'A is free'). For had the former claim been analysed in a way logically analogous to Day's analysis of 'A is unfree to D', we would arrive at the biconditional:

> A is free to D if, and only if, it is not the case that B makes A retrievably unable to D by Eing A.

And here, I think, we are more quick to see that what is being analysed is a person's being free to do or perform some act, but not 'free to' in a sense which is indistinguishable from claims regarding a person's freedom or liberty. That is, what we have is an analysis pertinent to 'free to' in the sense in which, for example, a person is free to return borrowed books to the library subsequent to the librarian having opened the doors for business; but not, I believe, an analysis which entails true (or false) claims about that person's liberty. Thus it would appear, if my objection holds, that there is more to liberty than our not having been made retrievably unable to do something by the interference of another.

Solano College © GEORGE E. PANICHAS 1979

THE UNEXPECTED EXAMINATION

By A. K. AUSTIN

IN a school there are two teachers, D and W. One Friday D tells his class that there will be an exam during the following week but that they will not be able to deduce, on the day before the exam, *that* it will be the next day. Also D explains that a pupil can deduce that p if and only if there is a set of propositions A such that the pupil knows or has been told every member of A and A entails that p.

The pupils in D's class immediately construct the following deductions.

Deduction 1. The exam will not be on Friday.

Reason: If it is then, on Thursday evening, we can deduce that it is on Friday. Contradiction.

Deduction 2. The exam will not be on Thursday.

Reason: If it is then, on Wednesday evening, we can deduce that it is on Thursday, using deduction 1. Contradiction. Similarly deductions 3, 4 and 5 for Wednesday, Tuesday and Monday respectively.

Deduction 6. The exam will be on Friday.

Reason: By deductions 2, 3, 4 and 5 it is not on any of the other days of the week. Similarly deductions 7 to 10 for the other days of the week.

Thus, on whatever day the exam is held, the pupils have, on the day before the exam, a deduction that it is on the next day. Hence D is unable to keep his word.

On another Friday W tells his class that there will be an exam during the following week but that they will not be able to deduce, on the day before the exam, on *which* of the remaining days it will be. Also W gives the following explanation. Let $p_1, \ldots, p_n$ be propositions. A pupil can deduce which of $p_1, \ldots, p_n$ is true if and only if, for exactly one i, he can deduce that p_i.

When Wednesday evening is reached without the exam being held, the pupils in W's class begin to make deductions. They deduce that it cannot be on Friday as they could deduce, on Thursday evening, on which of the remaining days it would be. Thus they have a deduction that it is on Thursday.

However, there remains the question of whether they can deduce that the exam is on Friday. Either they can or they cannot. If they can then, as they have already deduced that it is not on Friday, it follows that W's statements are inconsistent. On the other hand if they cannot deduce that it is on Friday then they are able to deduce which day it is. This contradicts W's statements and so shows they are inconsistent.

Thus W's statements are inconsistent and so any proposition can be deduced from them. Thus, except for the case when the exam is on Friday, the students are unable to deduce, on the day before the exam, on which day it will be.

On Thursday W holds the exam, so making his inconsistent statements come true.

University of Sheffield

NOTES

The ANALYSIS Committee consists of: Chairman, P. T. Geach; Secretary, J. H. Benson; Margaret A. Boden; A. E. Fisher, Andrew Harrison, R. F. Holland, Hidé Ishiguro, J. Kemp, Bernard Mayo, D. H. Mellor, R. G. Swinburne, A. R. White, C. J. F. Williams, Peter Winch. This committee is responsible for appointing and advising the Editor and for the general policy of the paper.

SUBSCRIPTIONS. The subscription to ANALYSIS for institutions is £5.00 (inland), £6.00 (overseas), $15.00 (U.S.A. and Canada); for individuals £4.00 (inland), £4.80 (overseas), $12.00 (U.S.A. and Canada). Each volume comprises four numbers, three of 48 pages and one of 64 pages, appearing within the academic year—in October, January, March and June. Orders should be sent to Basil Blackwell, 108 Cowley Road, Oxford OX4 1JF, or placed with any bookseller.

CONTRIBUTIONS. Articles submitted for publication should be addressed to Mr. Christopher Kirwan, Exeter College, Oxford OX1 3DP. Contributors are asked to note the following requirements.

Articles should **normally** not exceed 3,000 words in length. **Occasionally,** however, longer contributions can be accepted;

They must be type written in double spacing on one side of the paper only;

Footnotes should be kept to a minimum and wherever possible avoided altogether;

Single quotation-marks should normally be used, except for purposes of internal quotation and "scare" quotes.

Discussion papers should be sent in as soon as possible after the appearance of the article to which they refer.

It is regretted that owing to increased postal charges it is no longer possible to return typescripts unless the following instructions are followed:

Contributors in the United Kingdom should enclose a stamped addressed envelope of suitable size; if immediate acknowledgement is required, a stamped postcard should also be enclosed.

Overseas contributors who wish to have their MSS. returned should send an envelope **and international reply coupons** of the requisite value, whether for air or surface mail.

Galley proofs of accepted articles will be sent to authors for correction, together with information about offprints. Typescripts will be retained by the Editor on the assumption that authors have kept their own copies.

The copyright of articles printed in ANALYSIS remains the property of the author, but contributors are strongly advised, in their own interest, to consult the Editor before consenting to the reprinting of their articles.

ANALYSIS

Edited by

CHRISTOPHER KIRWAN

CONTENTS

events such that if they occur at all they occur prior to t, S at t presumably can do nothing which would cause F to be false.

Perhaps an analogous case will make the point here more obvious. Suppose S's pulling the trigger at t is sufficient given the conditions for Jones' death at $t+n$. But then, given these same conditions, Jones' *not* dying at $t+n$ is sufficient for it not being the case that S pulls the trigger at t. In other words, were Jones not to die at $t+n$, then given these conditions the proposition that S pulled the trigger at t would have to be false. But although this subjunctive conditional is true, Jones' not dying at t (if it in fact obtains) would not be the cause of S's not pulling the trigger at t.

The moral here is that subjunctives of the form 'if a were the case, b would be the case' can be true even if b occurs prior to a. Accordingly, the subjunctive in Lamb's revised principle can be true although the events picked out by F occur prior to the time t at which S would be refraining from A. In other words, it is true that if S at t were to refrain from A, then some member of F would be false, even though F refers only to events occurring prior to t. Moreover, it is perfectly plausible to suppose this subjunctive is true even though S at t can do nothing to alter the past.

Lamb, however, seems to think otherwise. He claims that the revised principle of can entailment can be used in conjunction with another principle, what he calls 'the principle of preexistence inefficacy', to demonstrate the inadequacy of compatibilism. This second principle is formulated as follows:

> If T is a non-empty set of true propositions each to the effect that an event E occurs, where E ends before person S comes into existence, then S at no time can refrain from some action such that were he to refrain from it some member of T would not be the case (p. 24).

Lamb claims that rejecting this second principle 'requires believing that man has the ability . . . to change the past'. But given the above remarks, it is clear that this principle can be rejected without implying that persons have the ability to alter the past. For again, if the past events picked out by T are sufficient for S's now doing A, then S's not doing A is sufficient for the non-occurrence of at least some of these events. So, if S were not to do A, then some of these events would not have occurred. But this truth is just a recognition of the fact that these events are sufficient for S's doing A and not an assertion to the effect that S can control the past.

If we were to suppose that a person does A at t only if at no prior time does there exist a sufficient causal condition either for the person's doing A at t or for his refraining from doing A at t, then the above principle of preexistence inefficacy might be acceptable. But neither

Lamb nor VanInwagen claims to have shown that determinism is in this way false. *Perhaps* they would claim that because freedom and determinism, they believe, are incompatible, it is more reasonable to believe that determinism is false than to believe that humans are not free. But this does not help them with the problems discussed here, which are concerned with their arguments against the compatibility of freedom and determinism. Indeed, insofar as a principle like the principle of preexistence inefficacy is used in such arguments and insofar as the principle is plausible only if determinism is assumed to be false, such arguments become uninteresting.

The upshot here is this: As Lamb originally formulates it, the principle of can entailment, *if* accepted, could be used to argue against compatibilism. But independently of any assessment of compatibilist attempts to analyse 'can' clauses of the sort that appear in this principle, there is no good reason to accept it. Moreover, as Lamb recognizes, there seems to be at least one kind of case which forces even non-compatibilists to reject his original principle. But if in order to avoid this kind of problem, we reformulate the principle of can entailment in the way that Lamb does, the principle can no longer be used to argue that compatibilists are committed to the view that if humans are free, they are able to alter the past.

University of Notre Dame   © RICHARD FOLEY 1979

AN ADVERBIAL MEINONGIAN THEORY

By WILLIAM J. RAPAPORT

I. Introduction

A fundamental assumption of Alexius Meinong's Theory of Objects (1904) is the act-content-object (ACO) analysis of psychological experiences. I suggest that Meinong's theory *need not* be based on this analysis, but that an *adverbial* theory might suffice. I then defend the adverbial alternative against a recent objection raised by Roderick Chisholm, and conclude by presenting an apparently more serious objection based on a paradox discovered by Romane Clark.

II. An Adverbial Meinongian Theory

According to Meinong (1904), every psychological experience is "directed" towards something called its "object" (*Gegenstand*) (pp. 483ff). ('Object' is here used more in the sense of 'that which is aimed at' than 'individual thing' and is perhaps best thought of for the moment as elliptical for 'object of thought', where 'thought' is generic for 'psychological act'.) This is a version of Brentano's Thesis of Intentionality (Brentano (1874) p. 50). As modified by Twardowski, it developed into the ACO-theory, according to which every psychological experience is analysable into an "act" which is directed to an object external to the experience by means of a "content" internal to the experience (cf. Grossmann (1974), Ch. III).

This is based on the fundamental datum that every judgment or idea is a judgment or idea *of* something, where this is interpreted to mean that there is an act and an object of the act. But it seems equally plausible to interpret it to mean that there is an act which has a certain characteristic or which is "performed" or experienced in a certain manner. Thus interpreted, there would be no "pure" judgments or ideas: just as there is no "pure" colour, but only red, blue, etc., so there would be only, e.g. mountainlike ideas, ghostly ideas, etc. On the former interpretation, however, there *is* a pure act of, say, judging, in the sense that the act is distinguishable from the object.

Nevertheless, on the alternative theory I wish to consider (roughly, one conflating the content with the object), the act is an experience *of a certain kind* or is experienced *in a certain manner*; this seems sufficient as an explication of the phenomenon of "directedness". I here make no commitment to the truth of this alternative; I am only concerned to see whether a Meinongian theory of objects would be impossible *were* the alternative true. Since the "content" was defined as that part of the psychological experience which "directs" the act to its object, let us

call this the "act-content (AC) theory". The AC-theory, then, holds that all ideas etc. *are* "of" something, *in the sense that* they all have a content.

It may help in clarifying the distinction between these two theories to consider the adverbial theory of perception. According to this theory,

> (1) I am sensing a red sensation (or, I am sensing a red square)

is to be explicated, not as a dyadic relation of sensing holding between a subject (I) and an object (a red sensation or a red square), but in a subject-predicate form as:

> (1A) I am sensing redly (or, as Sellars would have it, I am sensing a-red-square-ly).

(Cf. Chisholm (1957), Ch. 8 (1966), p. 95f; and Jackson *et al.* (1975) and the references therein.)

The generalization of this move to the case of thinking was, perhaps, first suggested by Wilfrid Sellars (1969, esp. pp. 235ff) and recently objected to by Roderick Chisholm (1973). On such a theory,

> (2) I am thinking of Plato

would be 'construed as telling us, not about something which is related to me as being the *object* of my thought, but only about the *way* in which I happen to be thinking' (Chisholm (1973), p. 210), e.g. as

> (2A) I am thinking Plato-ly.

(Roughly, when one thinks Plato-ly, one's thought processes (be they mental or physical) "present" to the thinker properties and characteristics which, we are inclined to say in ordinary language, are (thought to be) had by Plato.)

The AC-theory, then, may be taken as a version of an adverbial theory of thinking. On this theory, there are no "pure" acts of, say, thinking or fearing, nor is there any need for independent "objects" such as unicorns or ghosts which one might think about or fear. There would only be unified acts-of-a-kind or acts-in-a-kind-of-manner, such as "ghostly fearing". But clearly we can abstract an act of thinking and, so to speak, an "object" (i.e., a content or manner) of the act, and this abstracting allows us still to have a Theory of Objects. The Thesis of Intentionality can be preserved by interpreting it to mean that every act has a "manner", i.e., a content.

One difficulty is that the content is so intimately tied to the act that no two contents are identical, just as, on the ACO-theory, every two distinct acts have distinct contents, whether the acts be of distinct types or merely experienced by different people or at different times. Hence, in the AC-theory, we must talk of "content-types" or, perhaps, of

universals (or properties) whose particulars (or instances) are the individual contents (or content-tokens). Now, just as the ACO-theory must distinguish between individual acts (or act-tokens) and kinds of acts (or act-types) without thereby requiring a fourth component (making it an AtypeAtokenCO-theory), so the AC-theory, which needs content-tokens and content-types, need not be thought of as a three-component ACtypeCtoken-theory.

Nevertheless, the AC-theory augmented by content-types is isomorphic to the ACO-theory. Instead of a theory of *objects* on this view, we would have a theory of "manners" or contents. Such a theory would contain versions of the key theses of Meinong's original theory (cf. Rapaport (1978)). For example, suppose I think of the golden mountain: since the golden mountain is golden, the object of my thought is golden. On the AC-theory, this could mean that I am thinking goldenly and mountainly, and, so, I am thinking goldenly. And by means of the content-type we can explain how it is possible for two people to think of the "same" thing: the contents of their thoughts are of the same kind; i.e., they are thinking in the same manner.

In Rapaport (1978), it is suggested that the ACO-theory needs to be augmented by a fourth component after all, viz., by the "actual" object (*if any*) corresponding to the Meinongian one. There is even stronger reason for thus augmenting the AC-theory: If I think, e.g., of Jimmy Carter, we can distinguish four items: myself (the thinker), the act (thinking), the content (Jimmy-Carter-ly), *and Carter himself* (the actual, physical object).

III. Chisholm's Objection

Roderick Chisholm (1973) has objected that interpreting (2) as (2A) renders invalid an argument-form which had been valid before. Consider, first,

(3) Jones thinks of a unicorn.

This is paraphrased adverbially by Chisholm as

(3A) Jones thinks unicornically,

which is supposed to do away with the putative reference to unicorns and to have only to do with Jones, his act of thinking, and the manner of his thinking.

Consider, next, this valid inference:

(4)   (i) Jones thinks of a unicorn.
      (ii) Jones thinks only of things that exist.
∴ (iii) There are unicorns (i.e., a unicorn is a thing that exists).

Upon Chisholm's adverbial paraphrase, this becomes the *invalid* inference:

    (4A)   (i) Jones thinks unicornically.
           (ii) Jones thinks only of things that exist.
    ∴(iii) There are unicorns.

Chisholm's point is that (3A) (=(4Ai)) must still have something to do with *unicorns* to preserve the validity of the paraphrased inference.

Perhaps so. But (4A) is not the complete paraphrase. To obtain that, (4ii) would have to be adverbially interpreted also, as (let's say)

    (4AiiA) Jones thinks only existentially,

i.e., in whatever manner Jones thinks, some actual thing corresponds to his manner of thinking. (This will be made more precise in Section IV; cf. n. 1.)

Now, if (4Aiii) follows from (4Ai) and (4AiiA), then Chisholm's objection fails; otherwise, it is upheld.

I think that it does follow. For consider this valid inference:

    (5)   (i) Jones thinks of Quine.
        (ii) Jones thinks only of things that exist.
    ∴(iii) Quine exists.

Adverbially paraphrased *in toto*, I suggest that the premisses would become something like:

    (5A)   (i) Jones thinks Quinely.
         (ii) Jones thinks only existentially.

Now, the conclusion which follows from these adverbial premisses is:

    (5Aiii) Jones's Quinely thinking is existentially thinking,

i.e., some actual thing corresponds to Jones's Quinely thinking. Now, if (5A) is valid, as it seems to be, then (5Aiii) must be an adverbial reading of (5iii); i.e., to say that Quine exists is to say that (at least some instances of) thinking Quinely is (or, are instances of) thinking existentially. (This is spelled out in n. 1 and defended in Rapaport (1978).) So, to say that unicorns exist is to say that (at least some instances of) thinking unicornically is (or, are instances of) thinking existentially. Thus, the complete adverbial paraphrase of (4) is not (4A), but

    (4B)   (i) Jones thinks unicornically.
         (ii) Jones thinks only existentially.
    ∴(iii) Jones's unicornically thinking is existentially thinking.

Since this inference is valid, the adverbial theory is upheld.

IV. Clark's Paradox

The AC-theory has several advantages, including fitting neatly into a broader philosophical scheme along with the adverbial theory of perception, so that we might speak more generally of an Adverbial Theory of Mental Phenomena. Unfortunately, a paradox discovered by Romane Clark in a revised, ACO-style, Meinongian theory (reported on in Rapaport (1978)) rears its ugly head here, too. One way of presenting it is to turn to an interpretation of the *augmented* AC- (or adverbial) Meinongian theory.

If theories such as Meinong's are to embody the structure of the nature of thinking and its relation to the world, then any adequate neurophysiological theory about the nature of thinking ought at least to be consistent with such a structure. Suppose, then, that when one thinks *m*'ly, some (mental or physical) event is occurring in the thinker, having characteristics X, Y, etc., which, in turn, can be correlated in some way with (the property of being) *m*. For example, X, Y, etc., might be replaceable by some description of a sequence of neuron firings.

Now let *m* be some manner of thinking, say, a neuron-firing sequence correlated with my thinking of Pegasus. Surely, there is some sort of relationship between *m* and the properties ordinarily attributed to Pegasus (else, why would *m* be an act of thinking "of Pegasus"?); call this "relation" R. So, e.g., *m*R(flying horse), *m*R(creature of Greek mythology). Surely, too, any act of thinking *m*'ly will itself have properties, e.g., being a sequence of 13 neuron firings, or lasting for 1 second; call the relationship of *m* to such properties, S. So, e.g., *m*S (being a sequence of 13 neuron firings), *m*S(lasting for 1 second). Intuitively, for some property F, *m*RF iff *m* "represents" F "to us", and *m*SF iff *m* "(is conveyed by an act of thinking which) exemplifies" F. (My use of 'relationship' and '"relation"' is not intended to beg the question of the logical status of R or S. E.g., if S turns out to be, say, exemplification, it may or may not be a *relation*. Cf. Rapaport (1978), in. 12.)[1]

Now, some thoughts, as we ordinarily say, are "about" other thoughts. So it seems plausible that *m* might "exemplify" all of the properties it "represents" or, still more plausibly, perhaps, *m* might *fail* to exemplify some property it represents. E.g., if I think "about" a

---

[1] Using these notions, (5A) becomes:

(i) Jones thinks *m*'ly & (*m*'ly)RQ

(ii) $\forall M$'ly (Jones thinks M'ly$\rightarrow \exists \alpha \forall F((M$'ly)$RF\rightarrow \alpha SF$)) (i.e., in whatever manner Jones thinks, some actual object α exemplifies all of the properties represented to Jones by his manner of thinking)

(3) $\exists \alpha \forall F((m$'ly)$RF\rightarrow \alpha SF$)   (i, Simp, ii, UI, MP)

(4) $\forall F((m$'ly)$RF\rightarrow \alpha_1 SF$)   (3, EI)

(5) (*m*'ly)R$Q\rightarrow \alpha_1 SQ$ (4, UI)

(6) $\exists \alpha(\alpha SQ)$ (i, Simp, 5, MP)

thought which lasted for 1 second, and if that thought itself lasted for 1 second, then mS(lasting for 1 second) and mR(lasting for 1 second), We may express these possibilities thus:

$$\forall F(mRF \rightarrow mSF)$$
$$\exists F(mRF \ \& \ \sim(mSF)),$$

and we might consider two corresponding properties:

$$\lambda x \forall F(xRF \rightarrow xSF)$$
$$\lambda x \ \exists F(xRF \ \& \ \sim(xSF)).$$

Call these P and $\overline{P}$ respectively. Finally, suppose that I think "of" (to return once again to the more idiomatic ACO-talk) an object with *only* the property $\overline{P}$. Let m be the manner of my act of thinking thusly; i.e., $mR\overline{P}$ (and m "represents" nothing else).

Assume that mSP. Then $\forall F(mRF \rightarrow mSF)$, and, so, $mS\overline{P}$, which, on a reasonable requirement of consistency for the S-mode of predication (viz., $\forall F(mSF \longleftrightarrow \sim(mSF))$) contradicts our assumption. So, $\sim(mSP)$.

Assume that $\sim(mSP)$. Then $\exists F(mRF \ \& \ \sim(mSF))$, and, so (because m "represents" *only* $\overline{P}$), $\sim(mS\overline{P})$. This, on our consistency requirement, entails that mSP, contradicting our assumption. So, mSP.

But, either mSP or $\sim(mSP)$. So, *both* $\sim(mSP)$ *and* mSP. This is the adverbial version of Clark's paradox. (Both Clark and the referee have suggested to me that the present paradox is reminiscent of Grelling's.) The adverbial theory appears to be inconsistent.

Clearly, there are many challengeable steps in the derivation of this form of the paradox. The *most* challengeable, it seems to me, is the assumption that because (to revert once more to ACO-talk) I am thinking "of" an object which is *only* $\overline{P}$, therefore m represents $\overline{P}$ *and nothing else*. Perhaps the ACO-talk is *too* misleading; in the ACO-version of a Meinongian theory, a complex property *could* be the sole property of a Meinongian object, and the object's having that property would *not* force it to have any properties which might follow logically from its having that complex property. This is repugnant to many philosophers, though it is useful for resolving various philosophical puzzles (of the Hesperus-Phosphorus type; cf. Castañeda (1972)). The present suggestion is that the adverbial (AC) version might *not* have this repugnant feature and, thus, might avoid the paradox.

I am unhappy with this for several reasons. First, the usefulness of the lack of logical entailment just mentioned seems to be missing from the adverbial theory if we drop the assumption I characterized as 'most challengeable'. Second, and more importantly, this is an ad hoc way of avoiding the paradox. The paradox was discovered in connection with the ACO-version of a Meinongian theory (cf. Rapaport (1978)), and

ANALYSIS "PROBLEM" NO. 18

THE eighteenth problem is set by Professor Jonathan Harrison of the University of Nottingham under the title 'Jocasta's Crime'. Compulsory reading: 'Dr. Who and the Philosophers', by Jonathan Harrison, *Proceedings of the Aristotelian Society*, supplementary volume 1971, pp. 1–24; 'The Paradoxes of Time Travel', by David Lewis, *American Philosophical Quarterly*, 1976, pp. 145–52.

> Miss Jocasta Jones was walking in a secluded part of a local wood when she was attracted by the barking of her dog to a metal object which looked for all the world like an extremely old and rusty deep freeze. She opened it and found inside it a man, alive, but frozen solid. With some help from Miss Jones the man, who called himself Dum, thawed out. He handed Miss Jones a book, which he said told one how to build a time machine and a deep freeze.
>
> Miss Jones, however, was a very feminine creature, and much more interested in men than in engineering. She fell deeply in love with Dum, and married him. After a decent interval they produced a baby whom, because he was the spitten image of his father, they forenamed Dee.
>
> When Dee reached maturity, he found the book, which had carefully been put away on a high shelf, out of the reach of children. Following the instructions in it, he built the machine and got inside, taking his father and the book with him, in case he should need some technical assistance on his journey. He pressed the starter button, and, though nothing unusual appeared to happen inside the machine, everything visible through its solitary porthole seemed to start moving in the opposite direction to what it had before, and much more rapidly. The sun rose in the West, and set in the East, backwards day succeeded backwards day, and reverse year followed upon reverse year. The journey was so long that Dee, who had underestimated the amount of food they would need, was reluctantly compelled to make use of his greater youth and strength in order to kill and eat Dum. Eventually Dee arrived at the date which was his preselected destination, and got out. His first act on alighting from the time machine had been to blow it up, and everything contained in it, including what little remained of Dum.
>
> Dee, however, had not been happy in his new environment. The guilt he felt for what he seemed to remember he had done produced a mild attack of paranoia, which was accentuated by the loneliness of his position. He had wandered forlornly about, obsessed with the thought, which even changing his name to Dum could not eradicate, that despite his having destroyed the evidence, the men he now lived with could not fail to discover and severely punish him for his unnatural action. To escape he built the deep freeze, and got inside. He was prudent enough to take with him the book, in case he needed to build himself another time machine or deep freeze at an earlier or later date.
>
> The next thing Dum remembered was being resuscitated by a Miss Jocasta Jones, who had been exercising her dog in the neighbourhood . . .
>
> Did Jocasta commit a logically possible crime?

· The word limit is 600 words. Entries should reach the Editor of ANALYSIS by 31 August 1979; they should not be sent to Professor Harrison. Entries will not be acknowledged or returned unless accompanied by stamps or international postage coupons. Contributors may submit entries under their own names or a pseudonym. Contributors must be under the age of thirty, or undergraduates or graduate students.

A report with any winning entries will be published in volume 40 of ANALYSIS. The ANALYSIS Committee has voted a sum of £50 which will be awarded as a prize if Professor Harrison finds a sufficiently deserving contribution.

COMMENTS ON NEWTON-SMITH

By W. V. QUINE

W. H. Newton-Smith's paper 'The underdetermination of theory by data' (*Proceedings of the Aristotelian Society*, supplementary volume 52, 1978, pp. 71–91) is on the whole a defence of my views, and for this I am grateful. Even so, I seem to have been misunderstood at points, and these I shall try to clear up.

He begins with a reference to 'Quine's notorious claim that as *all* theories are underdetermined, translation is indeterminate'. The locus cited is my little 1970 paper 'On the reasons for the indeterminacy of translation' (*Journal of Philosophy* 67, 1970, pp. 178–83). In the sixteen lines that he quotes from it I conjectured that physical theory, the global system of the world, is underdetermined, but not that every subordinate system was underdetermined. Perhaps he could derive the latter 'strong thesis', as he calls it, from my doctrine of indeterminacy of translation, but he does not claim to; on the contrary, as we see from the sentence quoted above, he has me inferring the indeterminacy of translation by instantiation of the 'strong thesis'.

The actual argument in my 1970 paper was more devious. I began by imagining two global physical theories both of which accommodate the same totality of "possible observations", in a sense of this otherwise alarming phrase that I there explicated, and then I argued that either of these physical theories could be ascribed equally validly to the natives whose language was undergoing radical translation.

Let me say by the way that I have lost my liking for this particular

argument for the indeterminacy of translation in the case where one of the two competing theories is our own; for in devising a manual of translation I would favour agreement, where I could, between the natives and myself regarding the truth of a sentence and its translation. This policy would favour ascribing our physics rather than its rival.

Moving to Quine on observation sentences, Newton-Smith writes that 'certainly the assumption that there is a difference in kind between observational propositions and theoretical propositions is dubious'. I agree, and I have claimed no difference in kind. On the contrary, I find it essential for a theory of evidence that many theoretical terms turn up in many observation sentences. But it transpires, as he pursues the subject, that he was not really concerned about a difference *in kind*; he was worrying about how to distinguish between observation sentences and others *at all*. Can he have overlooked the social criterion that I stressed in *Word and Object* and various other writings? My criterion of observationality is simply the readiness of present witnesses to agree in their verdicts on the sentence on each occasion. There is some scope for gradations here, but it seems not to matter much.

Given my own chronic neglect of relevant literature, it would ill become me to complain that Mr. Newton-Smith missed my 1975 paper 'On empirically equivalent systems of the world' (*Erkenntnis* 9, 1975, pp. 313–28). That, however, was where I came to grips with questions of underdetermination that I touched only lightly in the 1970 paper that he cites. It was in the 1975 paper, in particular, that I took up the question of equivocation on which Newton-Smith cites Dummett: the question how much apparent underdetermination can be put down to equivocation. I put down much more of it to equivocation than Newton-Smith, in his defence of my supposed position, was prepared to concede.

I argued in that same unseen paper that Poincaré's well known finite model of the universe, in which things shrink as they move off centre, is only a verbal variant of classical theory and not a genuine alternative. Newton-Smith presents an example that is philosophically on a par with Poincaré's: he sets the hypothesis of a cyclically repetitive world history over against that of a finite time dimension. He means thereby to illustrate underdetermination, but on my view the two hypotheses are two formulations of a single theory.

Harvard University   © W. V. QUINE 1979

CARTESIAN INSANITY

By ROBERT FAHRNKOPF

WHETHER or not Descartes is entitled, in the *First Meditation*, to make an epistemological distinction between madness and dreaming, there is no good reason to doubt that he actually does draw this distinction, despite Steven DeHaven's arguments to the contrary ('Descartes, Frankfurt and Madmen', ANALYSIS 38.2, March 1978). DeHaven's textual interpretation rests upon the view that the remark 'praeclare sane' is an ironic or sarcastic comment upon, and therefore a repudiation of, Descartes' immediately prior dismissal of the possibility that he might be insane. But this textual interpretation is questionable on at least three counts.

First, DeHaven's claim that the *First Meditation* can be read as a dialogue between the meditator and the plain man—which allows DeHaven to attribute the remark about madness to the plain man rather than to Descartes himself *qua* meditator—is not merely a 'slightly artificial' reading, as DeHaven admits, but a totally gratuitous one. While the *Meditation* does have a dialectical character, this consists of a progressive sharpening and qualification of Descartes' own views, and not of a blunt juxtaposition of opposing views cast in dialogue form. Of those views which are advanced only to be subsequently abandoned, e.g., that I can know that $2 + 3 = 5$ whether I am awake or dreaming, none could be said to belong to the 'plain man'. The *First Meditation* simply gives no hint of the presence of any such unsophisticated spokesman. Furthermore, on the sole occasion when Descartes does discuss a view obviously not his own, he takes great pains to make explicit his own disavowal of this position. Thus:

> . . . there will perhaps be some persons who would prefer to deny the existence of so powerful a God . . . Let us not oppose them for the moment, and let us concede according to their point of view that everything which I have stated here about God is fictitious . . . (Lafleur translation).

In light of the deliberateness of this disavowal, the assumption that Descartes is content to use merely an ironic comment to dissociate himself from the earlier remark about madness seems implausible.

Secondly, in the French translation of the *Meditations*, which Descartes himself amended and approved, there is not the slightest indication of irony at this point in the text. After considering, and ostensibly rejecting, the possibility of his being insane, Descartes says simply:

Nevertheless, I must remember that I am a man, and that consequently I am accustomed to sleep . . . (Toutefois j'ai ici à considérer que je suis homme, et par conséquent que j'ai coutume de dormir. . . .)

The French 'toutefois' straightforwardly serves to reinforce the previous rejection of madness in favour of dreaming, and it is a mystery why Descartes should have let this translation stand if he were indeed trying to achieve the effect DeHaven supposes.

Finally, even conceding the irony in the Latin text, the most plausible target of this irony is not Descartes' rejection of the possibility of madness; rather, the target is the argument of which this rejection is only one premise, and which in its entirety might be put as follows:

(1) At least some of my perceptual judgments are not subject to error, unless I am insane.

(2) I am not insane.

∴(3) At least some of my perceptual judgments are not subject to error.

When, after presenting this argument, Descartes then says:

A fine argument! As though I were not a man who habitually sleeps at night and has the same impressions (or even wilder ones) in sleep as these men do when awake . . . (Anscombe and Geach translation)

it is obvious that he regards the argument as unsound in virtue of the falsity of the first, and not the second, premise. He is saying, in other words, that to know that he sees his hand in front of him, it is not enough to know that he is sane; although he does know this, he must also know that he is awake.

It seems clear, then, that Descartes does reject the possibility that he is insane. But is he justified in so doing? DeHaven thinks not, at least if dreaming is allowed as a legitimate epistemological threat, since, for Descartes, 'dreaming and madness are states of the very same sort', both states being characterized by 'perceptually bizarre experiences and judgments'. Of course, these states occur under quite different conditions—one commonly and during sleep, and the other rarely and in wakefulness—and they have different physiological causes—presumably only in the case of madness is the agitation of animal spirits in the brain due to the presence of dark bilious vapours—but these differences cannot serve to rule out the bare possibility of insanity. That Descartes does rule out this possibility shows that he regards insanity as something more serious than the propensity to be taken in by hallucinations; insanity must involve the forfeiture of rationality as such. If this sort of insanity is by definition impossible to attribute to a being created in the image of God, as Descartes believes man to be, then so much the worse for Descartes' claim to consistency. The fact remains that he regards

insanity as an inability to reason. And since Descartes' meditational program could not even be attempted unless he were able to make reasoned judgments, we should not be surprised, as Frankfurt is not, that the possibility of insanity is abruptly dismissed.

COMPATIBILISM AND CONTROL OVER THE PAST

By RICHARD FOLEY

ONE common way of arguing against the view that human freedom is compatible with determinism is to try to show that any such compatibilist position will be committed to the view that a person acts freely only if he has control over the past. And indeed, any argument sufficient to establish this would be sufficient to establish the unacceptability of compatibilism, since it is obviously implausible to claim that persons have control over the past. But, what sort of argument would be capable of showing that compatibilist positions do require free persons to have control over the past?

James W. Lamb and Peter VanInwagen are two philosophers who have recently tried to construct such an argument.[1] Lamb, for instance, claims that the required argument can be developed by appealing to what he calls 'a principle of can entailment'. Lamb's original formulation of this principle (his revised formulation will be discussed later) is as follows:

> If a set F of true propositions logically entails that S does action A then if S at time t can refrain from A there is some member of F such that at t S can make it false (p. 23).

Lamb does not try explicitly to defend this principle, although he does say that it 'verge(s) on being tautological'. But, VanInwagen, who offers an argument against compatibilism which is very similar to Lamb's, offers the following remarks in its defence:

> This principle seems to be analytic. For, if Q entails R, then the denial of R entails the denial of Q. Thus, any condition sufficient for the falsity of R is also sufficient for the falsity of Q. Therefore, if there is some condition that S can produce that is sufficient for the falsity of R, there is some condition (that same condition) that S can produce for the falsity of Q (p. 192).

[1] James W. Lamb, 'On a Proof of Incompatibilism', *The Philosophical Review* LXXXVI (1977), pp. 20–35; Peter VanInwagen, 'The Incompatibility of Free Will and Determinism', *Philosophical Studies* 27 (1975), pp. 185–199.

Now suppose that such a principle of can entailment is accepted and suppose also that determinism is understood to imply that for any action *a* which a person *S* might perform at time *t*, there is some set of conditions *c* that obtained at some prior time *t–n* such that the conjunction of *c* with certain laws of nature *l* entails that *S* does *a* at *t*. Given these two suppositions, it seems fairly simple to argue that compatibilism must be false. For, a compatibilist claims that although *c* and *l* together entail that *S* does *a* at *t*, if *S* could have done otherwise (in some appropriate sense of 'could'), *S* has acted freely. But given the principle of can entailment, *S* could have done otherwise only if either he at *t* could have changed the past or he at *t* could have acted contrary to laws of nature *l*. But then, if we assume that compatibilists would not want to claim that *S* here is free only if he is able to act contrary to laws *l*, compatibilists are forced to agree that a person like *S* acts freely at a given time *t* only if he at *t* is able to alter the past. However, this claim is obviously absurd. And hence, compatibilism must be abandoned.

But why, one might ask, should a compatibilist be inclined to accept the principle of can entailment? Even if he previously saw no reason to reject it, he should surely regard the above implications as a sufficient reason for its rejection. It is true that VanInwagen claims that the principle 'seems analytic' and Lamb says it 'verges on being tautological', but such claims are unconvincing unless they are accompanied by some suggestion about how to analyse the 'can' clauses in the principle. Indeed, if we grant that some compatibilist analysis of 'can' and 'could' statements is adequate, it is easy enough to turn Lamb's and VanInwagen's argument on its head and construct an argument against the principle of can entailment. This might be done, for example, by supposing that at least some claims of the form '*S* at *t* brings about *a* but he at *t* could have brought about *b*' are true and by further supposing that such claims are to be understood in the following compatibilist manner: '*S* at *t* brings about *a* but if he at *t* had willed to bring about *b*, this willing would have caused *b*'. Now, given also the truth of determinism, there must obtain some conditions *c* at a prior time *t–n* such that *c* in conjunction with laws of nature *l* entails that *S* at *t* brings about *a*. But if it is impossible for a willing at *t* to cause an event at a time prior to *t*, then it is false that *S* could have brought about not-*c* at *t–n*. And if we assume that it is false that *S* at *t* could have brought it about that laws of nature *l* were not in force between *t–n* and *t*, then the principle of can entailment must be false.

Of course, I do not endorse the above analysis of 'could have done otherwise'.[1] I have used it only to show how compatibilistic analyses of 'can' and 'could' statements might be used to argue against the principle

---

[1] For what I take to be a more plausible analysis of 'could have done otherwise', see Richard Foley, 'Compatibilism', *Mind* LXXXVII (1978), pp. 421–8.

of can entailment and accordingly to show why Lamb and VanInwagen, if they wish their arguments to be convincing, must demonstrate the inadequacy of such analyses. But this neither Lamb nor VanInwagen has done. (Moreover, *if* plausible arguments against compatibilistic analyses of 'can' and 'could' statements were available, then arguments using the principle of can entailment would seem to be unneeded. And so, it would also seem that Lamb's and VanInwagen's argument *at best* is superfluous.)

Such a response is sufficient in itself to blunt whatever force an incompatibilist argument based on the principle of can entailment might have. But, there is an additional consideration to be brought against the principle of can entailment. Indeed, Lamb recognizes this problem and alters his formulation of the principle to avoid it. The problem, in Lamb's words, is as follows:

> Let . . . F be the conjunctive proposition that T does B at 2:00 and S does A at 3:00. Let us suppose that this proposition is true and that at 3:00 S can refrain from A. According to the principle [of can-entailment], then, S should be able at 3:00 to make F false; indeed all he has to do, it seems, is to refrain from doing A. Suppose, however, that T is somehow able to find out before he himself does B whether or not S is going to do A. If he discovers that S will do A then he does B at 2:00; if, on the other hand, he discovers that S will not do A he refrains from doing B. It therefore seems to be the case that even though S can at 3:00 refrain from A he is unable at that time to make F false; were he to refrain from A, proposition F would have already been made false by T an hour earlier (p. 23).

In recognition of this problem with his original formulation of the principle of can entailment, Lamb alters the principle as follows:

> If a set F of true propositions logically entails that S does action A, then if S at time t can refrain from A then S at t can refrain from some action such that were he to refrain from it some member of F would be false (p. 23).

Lamb is correct in thinking that this revised principle avoids the kind of problem mentioned above. For if our person S were to refrain from A, some member of F would in fact be false, since T would have already refrained from B. The problem is, however, that the amended principle can no longer be used to argue against compatibilism. Indeed, the compatibilist, it seems, must accept this principle of can entailment. For if the truth of F entails that S does A at t, then the truth of F is sufficient for the truth of the proposition that S does A at t. But then, the falsity of the proposition that S does A at t is likewise sufficient for the falsity of F. So, if S were to refrain from doing A at t (and thus were to bring about the falsity of the proposition that he does A at t), F would have to be false. But the truth of this subjunctive conditional does not imply that S can bring about the falsity of F. For if F refers exclusively to

preliminary investigation suggests that it is also applicable to Frege's theory of sense and reference. It seems reasonable to require, then, that a *general* way out of the paradox be sought.[1]

State University of New York, College at Fredonia
© WILLIAM J. RAPAPORT 1979

REFERENCES

Franz Brentano, 'The Distinction between Mental and Physical Phenomena' (1874), trans. by D. B. Terrell, in *Realism and the Background of Phenomenology*, ed. by R. M. Chisholm (New York: Free Press, 1960): 39–61.
Hector-Neri Castañeda, 'Thinking and the Structure of the World' (1972), *Philosophia* 4 (1974): 3–40.
Roderick M. Chisholm, *Perceiving* (Ithaca, NY: Cornell University Press, 1957).
—— *Theory of Knowledge* (Englewood Cliffs, NJ: Prentice-Hall, 1966).
—— 'Homeless Objects', *Revue Internationale de Philosophie* 27 (1973): 207–23.
Reinhardt Grossmann, *Meinong* (London: Routledge and Kegan Paul, 1974).
Frank Jackson, Michael Tye, and Wilfred Sellars, 'Symposium: The Adverbial Theory of Perception', *Metaphilosophy* 6 (1975): 127–60.
Alexius Meinong, 'Über Gegenstandstheorie' (1904), in *Alexius Meinong Gesamtausgabe*, Vol. II, ed. by R. Haller (Graz, Austria: Akademische Druck- u. Verlagsanstalt, 1971): 481–535.
William J. Rapaport, 'Meinongian Theories and a Russellian Paradox', *Nous* 12 (1978).
Wilfred Sellars, 'Metaphysics and the Concept of a Person', in *The Logical Way of Doing Things*, ed. by K. Lambert (New Haven: Yale University Press, 1969): 219–52.

THE MODAL CONFUSION IN RAWLS'
ORIGINAL POSITION

By Michael E. Levin *and* Margarita Levin

IN a recent paper,[1] the second-named author has suggested that
Rawls' derivation of his two principles of justice from the "original
position"[2] is unsound because it neglects a certain distinction. The
distinction is that between a subject's not having a trait, and a subject's
having that trait but not knowing about it. Rawls supposes that a
"veil of ignorance" screens the participants in the original position[3]
from their personalities and abilities; but he vacillates between the
first and second way of regarding these hidden traits. We will not
defend here the claim that Rawls has in fact missed this distinction,
nor—what should at least be intuitively plausible—that recognizing
this distinction is critical for evaluating Rawls' derivation. Both these
points are expounded at length in PK. We will, rather, suggest an
explanation of *why* Rawls has missed this distinction—which, once
noted, seems so elementary. We will also suggest that even if the
source of this confusion is recognized, Rawls cannot consistently select
one or the other view. The distinction is based on two incompatible
ways of regarding the original position, and *both* ways of looking at
it play crucial roles in Rawls' argument.

Such explanations are not the least part of criticizing Rawls or
any philosopher: as Aristotle remarks (*Nicomachean Ethics*, 1154a 22–4)
you have not completely disposed of a philosopher's position until
you can explain why he made the mistake you attribute to him. Without
a plausible explanation for the commission of a purported error, the
suspicion will persist that one's opponent believes what he does because
he has noticed that it is true. By explaining his belief that *p* without
invoking *p*, we strengthen the hypothesis that *p* is false by showing
that a hypothesis which entails *p* is unnecessary. As we have already
suggested, however, our argument is not merely contingent upon
accepting the point made in PK. Our main purpose is not just
explanatory; the discussion presented below can be taken as an
independent argument against Rawls.

The "original position" is a possible world, in fact a garden-variety
possible world as philosophers have come to understand that notion in
the last decade. It is a way things could be, a possible situation, a possible
state of the world. Our suggestion, briefly, is that what has misled Rawls
is his vacillation between two inconsistent accounts of possible worlds.

---

[1] 'The Problem of Knowledge in the Original Position', *Auslegung* V (May 1978);
henceforth PK.
[2] John Rawls, *A Theory of Justice* (Harvard: University Press, 1971).
[3] Henceforth simply 'the participants'.

One is the *realist* view associated with David Lewis,[1] according to which possible worlds are just as real as the actual one. The other view, associated with Kripke,[2] is that the nature of a possible world is completely given by our stipulations about it. We may call the latter the *stipulative* view. What is important for our purposes is this one essential difference between them: according to the stipulative view, *all there is* to a possible world is what we say about it, while according to Lewis' view, a possible world—and the possible entities in that possible world— can and typically do have traits independent of those by which we specify or identify that world and its objects. Thus, suppose we describe a possible world w as one in which sentences $S_1, \ldots, S_n$ are true, and suppose S is independent of S_1 & ... & S_n. Then, on the stipulative view, S does not yet have a truth value in w. For S to receive a truth value in w, a further specification is required. According to realism, however, S does have a truth value in w—although we may not know what it is. To give an example, suppose I say 'What if Columbus had drowned on his way to the New World?' I am imagining a possible world w in which Columbus drowned. Now, for the stipulativist, the question 'What was Columbus wearing in w when he drowned?' has no answer—unless I go on to stipulate what he was wearing by expanding my supposition to (e.g.) 'What if Columbus had drowned on his way to the New World while wearing lederhosen?' Now for the realist I am, in the original sentence, talking about a possible world which does satisfy a maximal consistent set of sentences and in which, accordingly, Columbus was wearing something (although I don't know what). The realist might protest that my stipulation has not singled out any one possible world, since there are many possible worlds satisfying my description, in each of which Columbus is wearing something different. But still the realist holds that Columbus is already wearing something in each world of the set of worlds I have subtended, and the question 'What was Columbus wearing when he drowned?' is still well posed even if ambiguous.

It is not our business to decide between accounts of possibilia (although it seems to us that Kripke's account accords better with how we actually talk, and avoids paradoxes that beset realism). Rather, we want to suggest that Rawls—or, better, his argument—requires the original position to be taken sometimes realistically, sometimes stipulatively. Note, to begin with, that Rawls himself is entirely unclear when he addresses the status of the participants conceived as possible entities in a possible situation. He instructs us to 'keep in mind that the parties

---

[1] 'Counterpart Theory and Modal Logic', *Journal of Philosophy* (1967), and more fully expounded in *Counterfactuals* (Harvard: University Press, 1973) and 'Anselm and Actuality' in B. Brody, ed., *Readings in the Philosophy of Religion* (Prentice-Hall; New Jersey, 1974).

[2] See 'Naming and Necessity', in D. Davidson and G. Harman, *Semantics of Natural Language* (D. Reidel, 1972). Kripke therein characterizes realism as the idea that possible worlds are like distant galaxies seen through a telescope and known by description.

in the original position are theoretically defined individuals' (Rawls, p. 147). Does this mean they have *only* the properties the theory ascribes to them, even if the theory is not maximal consistent? This question is not answered by Rawls' remark, a few sentences later, that he aims to show that 'in the original position rational persons *so characterized* would make a certain decision' (ibid; our emphasis). The 'so characterized' operator is no more helpful than the '*qua*' in '*x qua A*'; Rawls never coherently explains its force (see PK). This state of affairs immediately leads to the confusion noted at the outset. Rawls is free to take the participants realistically or stipulatively: when taken realistically, they have traits they do not know about (see PK, p. 151–3); when taken stipulatively they do not. In fact, the *reason* they lack any traits but those permitted them by the "constraints on the original position" is just that, in this phase of the argument, the original position is being taken stipulatively.

Pressure to adopt each of these inconsistent stances comes from two different directions. First, there is the pressure from the deduction of the principles themselves. According to Rawls 'to say that a certain conception of justice would be chosen in the original position is equivalent to saying that rational deliberation satisfying certain conditions and restrictions would reach a certain conclusion' (Rawls, p. 138). Since we will not be directly concerned with the constraints or the principles, we will abbreviate them simply as C and P. What Rawls really wants, as part of his project of "moral geometry" (Rawls, p. 121) is that C actually be shown to *entail* P, that the constraints logically lead to the principles in question. That being so, talk about how the participants will reason is just a picturesque way of talking about the logical connections between C and P. The "constraints" are transcriptions of moral premises—for instance, 'the participants do not know their own names' is the transcription of 'principles of right contain no proper names'. It is as if one were to prove that a mathematical theorem T follows from axioms A by talking about how an ideal hypothetical reasoner, given only the information A and constrained by the rules of logic, would arrive at conclusion T. In fact, more difficult mathematical results are often presented in this heuristically valuable way. Proofs that one figure is constructible from another are sometimes couched in stories about how a geometer, armed with nothing but a description of the original figure, would go about constructing the desired one. In their textbook *Computability and Logic* Boolos and Jeffrey provide picturesque instructions for building the interpolated sentence in the proof of Craig's Interpolation Theorem. Now in all these cases a Kripkean view of the hypothetical mathematician is in order, for when we say 'imagine a mathematician trying to find the midpoint of a given line segment . . .' that is all we need to know about the mathematician.

Anyone who interjected 'but maybe he's hungry and gets distracted' has simply missed the point of the heuristic device. The Kripkean view is indeed required, for if we take the hypothetical mathematician to have traits beyond those specified in the story, such interjections are entirely in order (cf. Columbus' attire). In fact, if he is allowed to have traits beyond those specified by the hypothesis 'let x know that A' we have ruined the proof, for we then have no guarantee that his constructing the figure or finding the interpolant did not depend on some unspecified trait. The 'x' in our hypothesis must be a dummy variable (cf. the usual formulations of the rule of universal generalization in predicate calculus). If we take x to be anything more, the heuristic proof loses its mathematical rigour. In short: to the extent that the original position is a heuristic guide to the implications of moral premises transcribed as constraints—and Rawls often writes as if this is how we are supposed to take it—a Kripkean interpretation of the original position is required.

But the countervailing pressure is Rawls' insistence that we adopt what in PK is called the 'entry condition', the requirement that the original position should be 'interpreted so that one can at any time adopt its perspective' (Rawls, p. 139). Rawls stresses this condition because his argument is at bottom an argument from hypothetical consent; we *should* adopt P because certain suitably situated beings *would* adopt P (see PK, p. 149). Obviously, a hypothetical consent argument is stronger the more the hypothetical consenters resemble us (if only because their similarity to us suggests that *we* would consent if similarly situated); and, obviously, a hypothetical consent argument has no force if the hypothetical consenters are logically incomplete beings. So, to satisfy the entry condition, Rawls must think of the original position as a position *we* could be in, and, more important, of ourselves as potential participants. If this is so, Rawls' description of the original position does not exhaust the statements true in it; for to imagine that *I* am in the original position, and blinded by a veil of ignorance, is to imagine far more than that I *am* just what I am allowed to know. I *am* much more than that, unlike the hypothetical mathematician who just *is* a being who knows that a given geometrical object is a line segment. In short, the entry condition demands a realist view of the original position. But now, reasoning about what I would do—in particular, reasoning about the sort of reasoning I would find compelling —must take into account more than those facts about myself that I myself know. The reasoning required is of the sort characteristic of game theory construed as a branch of theoretical psychology; and a game-theoretical proof that I would follow a certain line of reasoning (e.g. maximin) must include, or at least be compatible with, my having those traits of mine that I am ignorant of. Perhaps the premises C I

am allowed to use do entail P; but—since we are now taking the original position realistically—it by no means follows that I will conclude P. Like the hypothetical mathematician realistically understood, I may be influenced by an inner state.

Conversely, a proof that I would adopt certain principles P would no longer be decodable as a proof that C entails P, since I am now more than an abstract embodiment of the premises C. In fact, some of the most critical junctures in Rawls' argument must be taken this way. Consider the idea that I and everyone else will be infinitely averse to risk and hence adopt the maximin strategy for choice under uncertainty. I do not know whether or not I would be infinitely averse to risk behind Rawls' veil; perhaps some powerful argument can be mounted to show that I would be. But so far as the formal constraints themselves go, the participants' aversion to risk is *completely unspecified*; as unspecified as the aversion of our hypothetical mathematician to any risk of failure to bisect the line. And a simple explanation for Rawls' failure to see this shift in his style of argument is that he has failed to notice his shift from a Kripkean to a realist conception of the original position as a hypothetical state of affairs.

In addition to explaining Rawls' vacillation between one's not having a trait and one's not knowing that one has it, the supposition that Rawls is simultaneously thinking of the original position stipulatively and realistically explains another odd feature of his argument. It has been remarked (most notably by Hare[1]) that the participants in the original position are curiously alike, and hence that having more than one of them is unnecessary. They all have the same knowledge and labour under the same ignorance, and all differences between them are assumed inert. All are equally (infinitely) averse to risk. They are evidently indiscernible —so why does Rawls have more than one of them, and why does he speak of their deliberations, coalitions, and the like? Why is Rawls oblivious to the fact that he is multiplying qualitatively indiscernible entities? The answer, very simply, is that, when Rawls is thinking realistically, he quite rightly finds them discernible. In themselves, beyond the stipulations of the original position, they *are* different— they have different "rational plans of life", desires, talents and abilities. They are indiscernible only with respect to their description in the original position: up to isomorphism, as it were. Of course—and here is the problem—these differences are allowed to play no role at all in the derivation of what they do and think because, once the derivation gets under way, Rawls must switch gears and imagine all his participants as *completely* described by the constraints. And, no doubt, readers who have to figure out the ground-rules of the derivation as they go along simi-

[1] R. M. Hare, 'Rawls' Theory of Justice', in N. Daniels, ed., *Reading Rawls* (Basic Books: N.Y., 1975).

larly seesaw between taking the participants as mere heuristic embodiments of *C*, and as actual people. This vacillation helps perpetuate the illusion that there is more than one participant, even though all the participants have all the same properties.

Readers sympathetic to Rawls might at this point say that, in spite of this confusion between stipulatively and realistically construed participants, *something* about the original position rings true. The persistence of this conviction derives, we think, from the great intuitive appeal of the following two ideas: (i) each of us *ought* to do what his moral "better half" would do, reason as his moral better half would reason; and (ii) the constraints on the original position capture, nearly enough, how our better half would reason. Even more compactly: we should do what our better halves would do. Rawls has shown what our better halves would do. Therefore, we should do that. But we must remember that this talk of "better halves" is metaphorical. "Better halves" come attached to whole actual people. Idea (ii) is about our "better halves" as stand-ins for a formal derivation; in (i) they must be *us* in a game-theoretical argument. This would not be so disastrous if the demonstration corresponding to (ii) were self-contained and formed the basis for the game-theoretical argument corresponding to (i). But in fact it is the game-theoretical argument that is supposed to justify the claim that the principles can be formally derived from the constraints! "Better halves" are not thin but autonomous entities to reason about, as Rawls in effect takes them to be in his Kripkean episodes. Rawls needs and promises[1] a proof that *we* would reason a certain way in certain conditions. The impression that Rawls has mounted an *argument* is sustained by taking one or the other derivation as fundamental in turn. And the reader (and evidently Rawls himself) does this by alternating between two incompatible pictures of modality.

Many critics attack Rawls by questioning his use of such technical ideas as maximin, indifference curves, and Pareto optimality. This approach seems to us to lop off branches rather than attack the problem at the root. (Surely the fundamental nature of justice cannot depend on whether social mobility is a stochastic process!: see Rawls, pp. 170f.) Such critics concede to Rawls almost everything of philosophical significance. If this paper and PK are right, *A Theory of Justice* is a non-starter because its crucial concept—the original position—rests on contradictory interpretations of the modal notions.

The City University of New York
© MICHAEL E. LEVIN *and* MARGARITA LEVIN 1979

---

[1] 'Thus it may be helpful to observe that one or more persons can at any time enter this position, or perhaps, better, simulate the deliberations of this hypothetical situation, simply by reasoning in accordance with the appropriate restrictions' (Rawls, p. 138).

ARISTOTLE AND TEMPORALLY RELATIVE MODALITIES

By MICHAEL J. WHITE

AS Professor Jaakko Hintikka has pointed out,[1] a temporal conception of the alethic modalities is repeatedly enunciated by Aristotle. An especially explicit statement of Aristotle's temporal conception of necessity is found in *De Gen. et Corr.* 2.11:

> For what is of necessity coincides with what is always, since that which must be cannot possibly not-be. Hence, a thing is eternal if its being is necessary: and if it is eternal, its being is necessary. And if, therefore, the coming-to-be of a thing is necessary, its coming-to-be is eternal; and if eternal, necessary. (337b35–338a4, adapted from H. H. Joachim's translation.)

Necessity becomes equivalent to eternal truth if this conception of necessity is cast in the formal mode: a proposition is necessary if and only if it is eternally true. Furthermore, if the 'standard' relations among the alethic modalities are accepted—as Aristotle generally seems to do[2]—it follows that a possible proposition is one true at some (at least one) time, a nonnecessary proposition is one that is not true at all times, an impossible proposition is one that is never true, and a contingent proposition is one that is true at some times but not at others.

If we confine our attention to *temporally indeterminate* or *unqualified* propositions, the temporal conception of the modalities does not seem implausible. We can think of such propositions (e.g., 'It is dark', 'There are Medes in the temple', 'Socrates will die') as propositional functions containing a free time variable; in normal use the value for the variable is supplied by the indexical 'present time'. There is some intuitive force to the supposition that such propositions (really, propositional functions) as '(Either it is day or it is not day) (t)' or '(All humans have kidneys) (t)' are true for all values of the variable and, thus, are necessary —or that the propositional function '(There exists a triangle with four sides) (t)', which is true for no value of the variable, is thus impossible. There is even some force to the claim that the 'sometime-truth/sometime-falsity' of temporally indeterminate propositions such as '(It is dark)(t)' or '(There are Medes in the temple) (t)' catches something of the idea of contingency; such propositions *can* be either true or false.

However, as Hintikka has remarked, the temporal conception of the modalities 'is bound to land [Aristotle] into trouble'[3] as soon as he

---

[1] See especially 'The Once and Future Sea Fight: Aristotle's Discussion of Future Contingents in *De ·Interpretatione* 9', reprinted in *Time and Necessity: Studies in Aristotle's Theory of Modality* (Oxford, 1973), pp. 147–178.

[2] *De Int.* 13.22a20 and 22b22; *An. Pr.* 1.13.32a25. Aristotle does, however, distinguish another sense of 'possibility', one that implies nonnecessity: *De. Int.* 13.22a15.

[3] *Op. cit.*, p. 151.

begins to consider *temporally determinate, qualified,* or *'fixed'* propositions. Such propositions can be thought of as not having a free time variable but as being bound or 'fixed' to a certain time. For example, consider the following sentence:

(1) Socrates is alive at 400 B.C.

If we make the most common assumptions about time, the sentence expresses an eternally true proposition. Had it been false, however, it would have expressed an eternally false proposition. If we adopt Aristotle's account of the modalities and add an extra (really superfluous) time variable to any such temporally determinate proposition (e.g., '(Socrates is alive at 400 B.C.) (t)'), it will be necessary if true or impossible if false. Thus temporally determinate or qualified propositions plus Aristotle's temporal semantics for the alethic modalities give rise to a form of fatalism or 'logical determinism': the mere truth of each member of a certain class of non-tautologous propositions, viz., the class of temporally determinate propositions, is sufficient to guarantee the necessity of the proposition in question.

However, my illustration of a temporally determinate proposition was obtained as a result of the occurrence of a calendar date in the sentence expressing the proposition. The illustration obviously assumes what Rescher and Urquhart would term a 'chronologically stable' time metric.[1] Yet, in the passage (*De Interpretatione* 9) in which Aristotle is most exercised with the problem of fatalism, no such examples occur. What occur, if Aristotle's examples are again cast foursquarely in the formal mode, are illustrations utilizing temporal indexicals (morphemes, especially time adverbs, the referents of which depend on the time of utterance of the constructions in which they occur). Most notorious is

(2) There will be a sea battle tomorrow.

I wish to suggest that such sentences as (2) are rendered importantly ambiguous by the temporal indexicals they contain. According to one reading of (2), the proposition it expresses is true on each day that precedes the day of a sea battle. Since there are some days on which sea battles do not occur, the proposition is not true on some days. Consequently, since it is not always true, the proposition is not necessary. This reading of the sentence does not, in fact, ascribe a temporally determinate proposition to it at all. Interestingly enough, there is evidence that such a reading was given this sentence by some ancient logicians. In the *De*

---

[1] Nicholas Rescher and Alasdair Urquhart, *Temporal Logic* (Vienna and New York, 1971), Ch. III.

Fato of Alexander of Aphrodisias, the opponent (the Stoic Cleanthes?)[1] argues that the proposition expressed by (2), even if in fact true, is not necessary, since the necessary is what is always true but this proposition will be false the day of the sea battle (presumably because it is assumed that there will not be a sea battle the day following *that* one, the day after tomorrow!).[2]

There is, however, another reading of (2) according to which it expresses a temporally determinate proposition. This reading utilizes what I shall call the rigidly referential or demonstrative sense of the temporal indexical 'tomorrow', herewith christened 'd-tomorrow' in honour of David Kaplan's 'dthat'.[3] This reading of (2) *forever* binds the proposition it expresses to the day after that picked out by the utterance of the sentence. Thus, if I use the sentence on 12 June, 1978, the proposition expressed is the same as that expressed by the following sentence (whenever it is employed):

(3) There is [in an atemporal sense] a sea battle on 13 June, 1978.

Of course, similar ambiguities occur in the case of sentences containing other temporal indexicals, e.g., 'now' ('today'), 'yesterday', 'two weeks hence', etc.

I suggest that it must be the temporally determinate ('d-tomorrow') readings of such sentences as (2) that concern Aristotle in *De Int.* 9. (Otherwise, he would be no more inclined to worry about such sentences expressing necessary propositions than the Stoic of Alexander's *De Fato* is.) Such propositions, if true, must be eternally true and thus necessary, according to Aristotle's temporal conception of the alethic modalities. It should be added that Aristotle seems quite willing to embrace logical determinism with respect to propositions bound to past (or present) times. If such a proposition is true, it is necessary; if false, it is impossible.[4] Aristotle's intuitive motivation here, I suppose, is

---

[1] *Supplementum Aristotelicum*, II, iii, ed. I. Bruns (Berlin, 1892), 10.177.7ff. The argument reported (and derided) by Alexander was perhaps a Stoic attempt to separate the concept of fate or destiny (εἱμαρμένη) from that of necessity (ἀνάγκη). The passage is rather obscure, but it is clear that Alexander thinks the argument is sophistical. The attribution of the argument to Cleanthes is due to P. M. Schuhl, *Le Dominateur et Les Possibles* (Paris, 1960), pp. 24–25.

[2] Alexander's basic objection seems to be that, although the Stoics are right in their claim that the proposition is 'not always true' and, thus, not necessary, this claim is irrelevant to the question of whether the sea fight originally picked out by the use of the proposition 'comes-to-be' of necessity (ἐξ ἀνάγκης).

[3] Kaplan's 'dthat' is a one-place functor, which, in his formal development of the logic of demonstratives, forms a constant from another constant. Its semantic force is to fix the referent of the original constant with respect to a context of use so that this referent remains the same 'in other worlds' (and 'at other times'). See David Kaplan, "The Logic of Demonstratives" (Mimeo, 1971, expanded version, 1973). Analogously, 'd-tomorrow' fixes the time referred to in the context of use so that it remains the same when the proposition is 'evaluated' at other times.

[4] Hintikka discusses the textual evidence in his paper "Aristotle and the 'Master Argument' of Diodorus", reprinted in *Time and Necessity*, p. 183.

that the past (and present) is now 'unchangeable'. The past is in no way alterable as we 'move into' the future; there just becomes more of it. However, Aristotle does not wish to embrace fatalism or logical determinism with respect to all propositions bound to future times. What is his solution?

Hintikka has elegantly argued that Aristotle basically shifts the discussion in *De Int* 9 back to temporally indeterminate or unqualified propositions. According to Hintikka, these are assumed by Aristotle to be the 'proper' bearers of the alethic modalities, at least insofar as discussions of fatalism or logical determinism are concerned. However, as Hintikka also admits, such a "solution" really leaves untouched the fact that 'all true statements about genuinely individual future events [as expressed, for example, in the 'd-tomorrow' reading of (2)] still remain necessary'.[1] Such a solution would seem merely to constitute an *ignoratio elenchi* to the genuine problem *De Int.* 9 raises with respect to fatalism and Aristotle's temporal conception of the modalities.

The "traditional" interpretation of *De Int.* 9, i.e. the view that part of Aristotle's solution to the problem of fatalism with respect to the future involves the denial of truth values to 'future contingent' propositions, at least has the merit, it seems to me, of being a philosophically more satisfying reading of the chapter. A temporally determinate proposition bound to a presently future time is indeterminate with respect to its truth value *until* we 'reach' that time, i.e., until that time 'becomes the present'. This much is commonplace in the secondary literature.

However, the introduction of (temporary) 'truth gaps' in this fashion must needs complicate Aristotle's temporal semantic conception of the alethic modalities. In particular, Aristotle would seem to need to relativize the modality of propositions to time in a way analogous to that in which he has apparently relativized the truth of (temporally determinate) propositions to time. For example, at one time (now), a temporally determinate proposition bound to a presently future time will be truth-valueless and thus, at this time (now), both possible and non-necessary; at a later time (when the time to which it is bound becomes present or past), it will become either necessary or impossible, according to Aristotle. Aristotle seems to recognize the need to relativize necessity to time in a passage emphasized (for a different reason) by Hintikka: 'for it is not the same thing [to say that] everything that is is of necessity *when it is* as [to say that] it is absolutely (ἁπλῶς) of necessity'. (19a26–28) Aristotle, of course, is never much more specific about the relativization of the modalities to time than the preceding passage suggests. However, bearing in mind the fact that Aristotle holds (a) that temporally determinate propositions bound to presently future times lack truth values (but gain them later) and (b) that propositions bound to present and

---

[1] Hintikka, "The Once and Future Sea Fight", in *Time and Necessity*, p. 159.

past times are either necessary or impossible (and remain so as we 'move into' the future), we can develop an 'Aristotelian' relativized temporal conception of the modalities.

Peter Woodruff's 3-valued logic can be adapted for this purpose.[1] '$p(t)$' represents a proposition (either temporally determinate or indeterminate) with a free time variable added. We know that if the proposition 'p' is temporally determinate and now true, it will be true for all future values of t; if it is temporally determinate and now false, it will be false for all future values of t. 't_0' represents the indexical present time, 'L' the necessity operator, and 'M' the possibility operator, defined in terms of the necessity operator in the usual way. The propositional connectives are Kleene's strong connectives. Most importantly, the matrix for negation is as follows:

$p(t)$	$\sim p(t)$
t	f
—	—
f	t

A one-place operator 'T' ('it is true that') is defined by the matrix

$p(t)$	$Tp(t)$
t	t
—	f
f	f

Two additional defined one-place operators are helpful:

$Fp(t)$ ['at t, it is false that p'] for $T \sim p(t)$
$Np(t)$ ['at t, it is neither true nor false that p'] for $\sim (Tp(t) \lor Fp(t))$

Now, the 'modified Aristotelian' alethic modalities can be set forth in terms of quantification over times.

1. $Lp(t_0)$ [at present, p is necessary]   iff $\forall t(t_0$ is not later than $t \supset Tp(t))$

2. $\sim Lp(t_0)$ iff $\exists t(t_0$ is not later than $t \cdot (Fp(t) \lor Np(t)))$

3. $L \sim p(t_0)$ [or $\sim Mp(t_0)$]   iff $\forall t(t_0$ is not later than $t \supset Fp(t))$

4. $Mp(t_0)$ [or $\sim L \sim p(t_0)$]   iff $\exists t(t_0$ is not later than $t \cdot (Tp(t) \lor Np(t)))$

[1] Peter W. Woodruff, 'Logic and Truth Value Gaps', in *Philosophical Problems in Logic: Some Recent Developments*, ed. Karel Lambert (Dordrecht, (1970), pp. 121–142. There are a number of ways to 'save' the theorems of classical, two-valued logic (including the so-called theorem of 'excluded middle', which Aristotle has been interpreted as accepting). One way is to define logical truth in terms of classical valuations in the manner of van Fraassen's supervaluations; another is to employ Woodruff's concept of 'weak' or 'hedged' assertion for this purpose.

It should be noted that, given the standard relation between '*L*' and '*M*' and the preceding matrices and definitions, the biconditionals 2–4 follow from 1.

This modification of Aristotle's nonrelativized temporal account of the modalities obviously involves considerable extrapolation from what he himself says. I do not have the space to attempt any sort of detailed textual defence of this extrapolation, and am uncertain that I could do so anyhow. Instead, I wish to conclude with two points. The first is that the proposed temporal relativization of Aristotle's conception of the modalities seems to reconcile the conflict between (a) his acceptance of a temporal semantic conception of the modalities and (b) his desire to avoid fatalism or logical determinism with respect to the future, a conflict created by *temporally determinate* propositions of the sort Aristotle seems concerned with in *De Int.* 9. The second point is that the temporal relativization of Aristotle's account of the modalities that I have set forth bears a striking resemblance to the account of the alethic modalities attributed to Aristotle's somewhat younger contemporary, the Megarian logician Diodorus Cronos.[1] The two accounts of necessity and impossibility are identical. The accounts of possibility differ only in that, for Diodorus, a proposition is presently possible if and only if it is true now or at some future time, while for 'Aristotle', a proposition is presently possible if and only if it is true *or* neuter now or at some future time. There is an analogous difference in the two accounts of nonnecessity. The upshot of the difference between them is that Diodorus' account entails that any true *temporally determinate* proposition is necessary and that any false *temporally determinate* proposition is impossible. This is so because Diodorus evidently does not allow truth value gaps; thus he, unlike Aristotle, must maintain that a temporally determinate proposition is either true at all times or false at all times, including those times that occur *before* the time to which the proposition is bound. He thus embraces the sort of logical determinism that I have argued Aristotle seeks to avoid by means of the doctrine of truth gaps that get filled in 'in the fullness of time'. Happily, this difference accords with what little we know about Diodorus' views. Cicero reports that Diodorus gladly accepted the fatalism with respect to the future that Aristotle takes pains to avoid.[2]

Arizona State University   © MICHAEL J. WHITE 1979

[1] Boethius, *Commentarii in Librum Aristotelis (In De Interpretatione)*, second edition, ed. C. Meiser (Leipzig, 1877), p. 234.
[2] Cicero, *De Fato* 7.13 (Cambridge, 1960).

THE PARATACTIC ACCOUNT OF SAYING OF

By S. D. GUTTENPLAN

I

DAVIDSON'S proposals for a paratactic account of the logical form of certain indirect discourse claims and his use of a *same-saying* relation to fill out the truth conditions of these claims have found wide but by no means universal acceptance. Other problems aside, a very great obstacle for his proposal has been its apparent inability to help us with the notoriously difficult 'saying of' idiom; an idiom which has intimate semantical (if not logical) connections to the 'saying that' idiom—the original subject of Davidson's account. In fact, in a more recent piece, Davidson has despaired of the possibility of simply extending his treatment of 'said that' to 'said of . . . that'.[1]

Not all of his supporters, however, share his pessimism. In 'Saying of' (ANALYSIS 37.4, 1977) Jennifer Hornsby claims to see a way to extend the original account to cover the recalcitrant idiom. In this paper I shall present an objection to this attempt which is, I think, sufficiently general to rule out all proposals along the lines sketched by Hornsby. First, though, two preliminary remarks need making; remarks which will allow me to present my objection without being drawn too deeply into the morass of examples surrounding attitude claims.

II

The example will be reassuringly familiar. Ralph says:

(1) 'Hesperus was visible yesterday.'

Consider three claims that might be made about Ralph's performance:

(2) Ralph said that Hesperus was visible yesterday.
(3) Ralph said that Phosphorus was visible yesterday.
(4) Ralph said of Phosphorus that it was visible yesterday.

Given the astronomical knowledge that philosophers have absorbed from examples such as this and given certain, by now traditional, views of these sentences, many would see (2) and (4) as true and (3) as false.

More recent work on attitude claims and singular terms, however, has made the above answers seem far too simple.[2] First, (3) can be heard in a way which makes it no less true than, and perhaps equivalent to, (4); speakers with no philosophical knowledge of these examples often

---

[1] In the introduction to *Logic and Grammar* edited by Davidson and Harman.
[2] I have in mind here Brian Loar's 'Reference and Propositional Attitudes' (*Philosophical Review*, 1972); Marc Temin's 'The Relational Sense of Indirect Discourse' (*Journal of Philosophy*, 1975); and Jennifer Hornsby's 'Singular Terms in Contexts of Propositional Attitude' (*Mind*, 1977).

hear (3) and (4) this way. In addition our intuitions about attitude sentences seem to change when singular terms other than proper names occur in these sentences. Both of these sophistications are important but neither will affect my objection to Hornsby's account and I propose, therefore, the following restrictions:

(i) the constructions in (2) and (3) are to be understood in a way that makes (2) true and (3) false. That is, we are to see them as *opaque* to substitutions for at least the singular terms following the 'that'. The construction in (4), however, is to be understood as transparent to such substitutions.

(ii) I shall only consider examples of indirect discourse where the original saying can be reported *truly* both by a sentence of the form of (2) *and* by a sentence of the form of (4); though in these sentences we should not expect to have the same singular terms.

The purpose of the first restriction is obvious enough. My interest is in a perfectly clear use of these two idioms though that use does not fit, at all points, the ordinary one. The hope is that the technical use can be employed to illuminate the ordinary one—to 'disambiguate' it if one chooses to see the ordinary use as ambiguous.[1] The footnote on p. 179 of Hornsby's paper show that for the purposes of her proposal she accepts this restriction to 'logician's English'. A further but crucial purpose of this restriction in the present paper is to forestall certain objections that might be made to my argument in section III.

The second restriction is made largely to keep my intended objection in sharp focus. Examples in Temin's and Loar's papers should make us doubt the availability of both of these idioms in all cases of reported speech and this because of the variety both of singular terms which reporter and reportee could use and of circumstances in which they are so used. However, Hornsby's paper proposes an extension of Davidson's account of 'said that' to 'said of . . . that' (both in logician's English) so I feel justified in treating only those cases where claims of both sorts are warranted.

III

Davidson claimed that the logical form of:

(2) Ralph said that Hesperus was visible yesterday.

is:

(5) Ralph said that. Hesperus was visible yesterday.

where 'that' is demonstrative of the utterance of the sentence following the full stop. Moreover, he maintained that we decide the truth of 'Ralph said that.' (as used above in (5)) by seeing whether the demon-

[1] Temin sees it this way. See p. 304 of his article cited in footnote 2.

strated sentence—a sentence used by the person making the claim—
samesays with some saying of Ralph's. Having given us the logical form of
(2) and a way of seeing how to decide its truth, Davidson felt that he had
done enough to solve the major problems posed by the 'said that'
idiom.

Hornsby uses Davidson's work on the logical form of action
sentences to give the following more detailed representation of (1):

(6) ($\exists u$) (Said (Ralph, u, that)). Hesperus was visible yesterday.

Insofar as 'Ralph said that.' is an action sentence this move is obviously
acceptable—barring reservations about the Davidsonian treatment of
action sentences. The importance of this otherwise small addition to (5)
is that (6) quite naturally allows adverbial modification in the con-
junctive style and this is precisely one of the ingredients Hornsby needs
for her treatment of:

(7) Ralph said of Hesperus that it was visible yesterday.

For, using the style of (6), she claims that (7) is to be understood as of
this form:

(8) ($\exists u$) (Said (Ralph, u, that) & Of (Hesperus, u)). x was visible
yesterday.

where the major emendation of Davidson's account is that 'that' now
demonstrates an utterance of an open sentence. That such demonstration
is coherent—and is even a part of English idiom as the first two sentences
in her article are meant to show—is not something I shall take issue with.

Given this change, however, Hornsby is obliged to provide us with
some idea of how we are to go about deciding the truth of (8); an idea,
that is, that plays the role of Davidson's *samesaying*. She does this by
generalizing Davidson's notion as follows:

> a closed utterance u SAMESAYS with an open sentence v if and only if
> u is of an object and what it predicates of that object is the same in import
> as v. (*op. cit.* p. 179)

My objection to her account arises from this and any similar proposed
generalization.

Let us go back briefly to Davidson's original use of the samesaying
relation—a relation he seems to have thought of as obtaining between
closed utterances. Ralph had uttered the sentence 'Hesperus was visible
yesterday' and someone, say John, had correctly reported this with (2).
Since, by Davidson's lights, (2) is of the form given in (5), nothing
prevents our using the identity we know to obtain to infer:

(9) Phosphorus was visible yesterday.

event, whereas an event does not. Or, for any complex c, there is an event e such that $(\forall t)$ (c happens at $t \to e$ occurs at t): for any event e, there is no other event e' such that $(\forall t)$ (e occurs at $t \to e'$ occurs at t).

Can complexes ever be causes or effects? I am not certain what to say about the possibility of their being causes. Nevertheless it seems to gain an intuitive warrant that they are eligible to be effects; that Ann's conception was a cause of her giving birth *and* a cause of John's becoming an uncle, that Socrates's drinking hemlock was a cause of his death *and* a cause of Xanthippe's becoming a widow, and that Bill's seeing a crossroad was a cause of his raising his arm *and* a cause of his signalling. Causing e to occur is one way of causing complex $\{e, s\}$ to happen. This intuition can be put as follows.

> Event e causes complex $c = \{e', s\}$ to happen at t if s holds at t and e causes e' to occur at t.

Event e causes complex c to happen iff e causes c to happen at some time. Obviously, this definition is in harmony with the revised counterfactual analysis of causation.[1]

Princeton University   © T. YAGISAWA 1979

A BETTER ALTERNATIVE

By MICHAEL PERLOFF

SINCE Bergström's original and provocative work a substantial debate has been continued concerning the nature of *alternatives to an action*.[1] That debate has been directed toward producing an account of alternatives which would make that notion both internally coherent and consistent with a reasonable form of act-utilitarianism. In a recent discussion Bergström has said that 'there is no inter-subjective technique for discovering relevant alternative-sets' ([4] p. 143). I think this plaintive remark to be too pessimistic. If it were true we, as actors in the world, would be severely limited in our understanding of ourselves and others; and we, as philosophers, would be severely limited in our understanding of alternatives.

Fortunately, for all of us, there is reason to believe that it is false. I shall try to show that given two quite reasonable assumptions, headway can be made toward an inter-subjective technique for discovering relevant alternative sets. The first assumption concerns the actions performable by an agent (either in a situation or at a time). The assumption is that there are some actions performable by an agent and that an acceptable list of them can be determined. We may want the list to be acceptable to the scientific community, or to the philosophic community, or to those concerned with the actor's future, or merely to the actor himself. However "acceptable" is understood, I assume that such a list may be compiled. Notice that if *no* list is acceptable, then no account of alternatives is possible. Also, it appears that we can and do know of ourselves and others what actions are performable.

The second assumption is that alternatives are best understood not as bare actions but as courses of action. That is, serious consideration of alternatives necessitates that we think of actions as parts of rival courses of action, for it is only within this wider context that actions may be properly judged and compared.

On these two assumptions it can be shown that

> For any acceptable list of actions performable by an actor, the alternative courses of action for the actor are determinable.

Assume an acceptable list of performable actions $a_1 \ldots \ldots a_n$. Since nothing is required of an action other than performability, it will make no difference whether these actions are particular, concrete, or versions of one another. Following Bergström we will call the set which contains only an action and its forbearance a minimal alternative set $\{a_i, -a_i\}$

---

[1] The seminal work is Bergström's [2]. Bibliography and an overview of the discussion can be found in [1] and [4].

([2] p. 42). For each action on the list there is a minimal alternative set, so there will be a family of minimal alternative sets. The cartesian product of this family of sets will contain various courses of action, some of which are performable, some impossible. The alternative courses of action are those sets whose members are jointly performable actions; and there is a subset of the cartesian product each member of which is such a course of action. This, then, is the relevant alternative set, a set of courses of action, where each course is a jointly performable set of actions.[1]

Let us take the following case in which the acceptable list is as follows:

a = Peter studies philosophy
a_1 = Peter studies philosophy at Uppsala
a_2 = Peter studies philosophy at Oxford.

The minimal alternative sets are $\{a, -a\}$ $\{a_1, -a_1\}$ $\{a_2, -a_2\}$. The cartesian product, or all the courses of action are

$$\{(a, a_1, a_2), (a, a_1, -a_2), (a, -a_1, a_2), (a, -a_1, -a_2), (-a, a_1, a_2),$$
$$(-a, a_1, -a_2), (-a, -a_1, a_2), (-a, -a_1, -a_2)\}$$

From this set the alternative courses of action comprise the following set:

$$\{(a, a_1, -a_2), (a, -a_1, a_2), (a, -a_1, -a_2), (-a, -a_1, -a_2)\}$$

Notice that among these alternatives there is one problematic set, $(a, -a_1, -a_2)$. That course of action is the one in which Peter studies philosophy but at neither Uppsala nor Oxford. It might be claimed that its inclusion shows that even given this procedure we cannot determine the alternative courses of action, for we don't know which subset is the correct one. It seems to me that the problem is not one of alternatives, but rather concerns performable actions. It is surely conceptually possible for Peter to study philosophy without attending either Uppsala or Oxford, therefore it should be among the alternatives. It might be argued that since Peter must study philosophy in some way the original list could not have been an acceptable one. However one decides this issue it illuminates an important feature of this account which is that problems of performability are separable from those of alternation. Whatever epistemological or ontological difficulties may arise concerning performability, or even the joint performability of courses of action, they may be held in abeyance while the study of alternatives continues.

Several features of this proposal are quite favourable. By focussing

[1] The solution presented here is a variation of an idea presented by Castañeda in [5]. I hope that this solution should satisfy Bergström's misgivings in [3].

on courses of action we produce an account sufficiently general to be compatible with various theories of action and differing accounts of moral obligation. Since every alternative course of action is incompatible with every other alternative course of action problems of inconsistent obligation do not arise. Also this account accords nicely with our actual practices in deliberation. In the course of practical reasoning we rehearse our repertoire of performable actions and then go on to imagine the different alternative courses of action open to us.

I take the next feature to be a very positive one, though I imagine that some philosophers will not. It is that the logical relations among actions are to be analysed as relations among courses of action. So to say that one ought to perform the conjunctive act a & b, is to say that one ought to perform the actions in a course of action which contains both a and b. That one ought not to perform b is analysed as claiming that any course of action undertaken by the agent ought to contain $-b$. In this way the problem of conjunctive actions is avoided, for it will never be true that an agent is obligated to perform some conjunctive act without being obligated to perform both conjuncts.

Finally we are in a position to state the utilitarian thesis

(T) An action ought to be performed if, and only if, when performed, that action is performed as part of a course of action whose consequences are better than those of every other alternative course of action.

University of Wisconsin-Milwaukee   © MICHAEL PERLOFF 1979

REFERENCES

[1] Bales, R. E., 'Review', *Theoria*, Vol. 40 (1974), pp. 35–57.
[2] Bergström, L., *The Alternatives and Consequences of Actions* (Stockholm: Almquist & Wiksell), 1966.
[3] —— 'On the Coherence of Act-Utilitarianism', ANALYSIS, 33.3 (1973), pp. 98–102.
[4] —— 'On the Formulation and Application of Utilitarianism,' *Noûs*, Vol. 10 (1976), pp. 121–144.
[5] Castañeda, H. N., 'A Problem for Utilitarianism', ANALYSIS, 28.4 (1968), pp. 141–142.
[6] —— 'Ought, Value, and Utilitarianism', *American Philosophical Quarterly*, Vol. 6 (1969), pp. 257–275.

IMPERATIVE INFERENCE

By VERA PEETZ

CAN we draw, deductively, an imperative conclusion from premises which are either all imperatives or a mixture of imperatives and indicatives (or even all indicatives, as both Professor Hare[1] and Professor Geach[2] have suggested)? Professor Bernard Williams[3] has expressed doubts about drawing such inferences, but there are many (for example, Professor Hare and Professor Rescher[4]) who not only think that imperative inference is possible but have put forward rules for, and have given definitions of validity in, imperative inference. A typical definition of validity in imperative inference is that an argument is valid if when the premises are obeyed or true (according as they are imperative or indicative) then necessarily the conclusion is also obeyed. This is parallel to the definition of validity in propositional inference that an argument is valid if when the premises are true, then necessarily the conclusion is also true. Difficulties might be raised as to whether compliance with an order (or orders) in the premises would ensure compliance with the order given in the conclusion, but I do not want to discuss such difficulties. Rather, I would like to ask whether an argument which satisfies such a definition of validity in imperative inference could be uttered in a speech context.

Let us consider Hare's well-known example

(A) (i) Take all the boxes to the station
      (ii) This is one of the boxes
      (iii) Therefore take this box to the station.

This argument satisfies the definition of validity given above and appears to be very similar to the following example of a valid propositional argument

(B) (iv) All men are mortal
     (v) Socrates is a man
     (vi) Therefore Socrates is mortal.

About both these arguments it can be said firstly that anyone uttering the premises is committed to the conclusion, in the sense that it would be inconsistent of him to utter the premises together with the negation of the conclusion, and secondly that, in a sense, the conclusion is contained in the premises.

Propositional logic is concerned with the logical relations between

---

[1] *Language of Morals*, O.U.P., 1964
[2] 'Imperative and Deontic Logic', ANALYSIS, 18.3, 1957–8.
[3] 'Imperative Inference', ANALYSIS, 23 supplement, 1963.
[4] *The Logic of Commands*, London and New York, 1966.

propositions, not with the illocutionary forces with which these propositions might be put forward in any actual speech situation. Similarly, in imperative logic, logical relations will hold between imperative locutions (or between imperative locutions and propositions) and will not be concerned with the forces with which they might be uttered. Or, as Professor Rescher puts it (*op. cit.*): 'exactly as standard (assertoric) logic occupies itself almost solely with the *content of the assertion* . . . so our logical theory of commands will dwell, in the abstract, upon the content of the *command given* or the *order issued* . . .'

Although argument (B) satisfies the definition of validity given, is it possible to assign illocutionary forces to the components of (B) so that we can utter (B) as an argument in ordinary discourse? In fact, we can assign such forces and the components of (B) made illocutionarily explicit will be (or will be something like)

(B')   (iv') I state that all men are mortal
        (v') I state that Socrates is a man
       (vi') I conclude that therefore Socrates is mortal.

Let us now look at Hare's argument (A). It seems clear that those who think that this is an example of a valid imperative inference, for example Hare and Rescher, would think that this example made illocutionarily explicit would be

(A')   (i') I order that you take all the boxes to the station
      (ii') I state that this is one of the boxes
      (iii') I order that therefore you take this box to the station.

If (A) is considered to be an argument which can be uttered in the context of normal discourse, then the conclusion must have the illocutionary force 'I conclude' and not 'I order'; for 'I order' gives no indication that a conclusion has been drawn. However, if the illocutionary force of the conclusion is to be 'I conclude', then the content of this act cannot be an imperative but must be a proposition, albeit a proposition describing an imperative. So we have

(A'')   (i'') I order that you take all the boxes to the station
       (ii'') I state that this is one of the boxes
       (iii'') I conclude that therefore you are (ordered) to take this box to the station.

In (iii''), 'therefore' indicates the grounds for the conclusion. But can a ground (or a reason) for a conclusion be an order? Grounds (or reasons) are grounds *that* or reasons *that* and so must be expressed as propositions; that is, the premises from which a conclusion is drawn must be propositions and cannot be imperatives. Another way of putting this point is to say that inferences can always be cast in the form 'since

p, *q*,' and imperatives cannot occur in 'since' clauses[1]. So we will now have

> (A‴)   (i‴) I state that you are (ordered) to take all the boxes to the station
>   (ii‴) I state that this is one of the boxes
>   (iii‴) I conclude that therefore you are (ordered) to take this box to the station.

Hare's non-explicit argument has now become

> (Aᴵⱽ)   (iᴵⱽ) You are (ordered) to take all the boxes to the station
>   (iiᴵⱽ) This is one of the boxes
>   (iiiᴵⱽ) Therefore you are (ordered) to take this box to the station.

This argument is a valid *propositional* argument.

Imperative inference is not possible, therefore, in the context of speech acts. The constituents of a deductive argument must be propositions.

Although we cannot draw deductive imperative inferences, we can and do draw imperative conclusions, although not imperative conclusions deductively arrived at. For example,

> (E) (vii) Your aunt is ill.
>   (viii) She was asking about you.
>   (ix) Therefore, go to see her.

This is not a deductive argument; (vii) and (viii) provide the reasons for ordering (ix). 'Therefore' plays a different rôle here from the rôle it plays in, for example, (iii‴), 'I conclude that therefore you are to take this box to the station'. In (iii‴), 'therefore' indicates the grounds for the conclusion. In (ix), 'Therefore go to see her', 'therefore' indicates the reasons for giving the order. In its fully explicit form (E) would be

> (E') (vii') I state that your aunt is ill
>   (viii') I state that she was asking about you
>   (ix') I order you to go to see her. I state that I am ordering you to do this because of the reasons given in (vii') & (viii').

It is important to distinguish between the two types of argument: arguments such as (E) and the deductive type of argument (Aᴵⱽ) discussed above. Whereas in the conclusion of (E) the speaker gives an order, in the conclusion of a deductive argument, even where it is a proposition describing an order, no order is given or endorsed by the speaker. As we have seen, Hare's argument (A) is not, as it stands, a

---

[1] I owe this point to the Editor, to whom I am grateful for criticisms of an earlier version of this paper.

deductive argument in the context of speech, but it could be regarded as an argument of the same type as (E) although, even here, the first premise will have to be transformed into a proposition (for (E) can be put in the form 'since p, do x' and an imperative cannot occur in a 'since' clause):

(F) (x) You are to take all the boxes to the station
  (xi) This is one of the boxes
  (xii) Therefore, take this box to the station.

(F) is not a deductive argument. The order (xii) is given for the reasons given in (x) and (xi).

This distinction is important for moral philosophers if they believe as Hare does, 'that it must be part of the function of a moral judgement to prescribe or guide choices' (*op. cit.* p. 29). In spite of what Hare says about imperative inference, no *deductive* imperative conclusion can be drawn from premises either imperative, indicative or both.

Hence if a moral judgement is to prescribe and if this prescription is to be the conclusion of an argument, the argument cannot be a deductive one.

Incidentally, the much-debated question of whether one can infer 'Post the letter or burn it' from 'Post the letter' can now be settled. Since one cannot make imperative deductions, one cannot infer from the imperative 'Post the letter' the imperative 'Post the letter or burn it'. All that one can do is from 'You are ordered to post the letter' to infer 'You are ordered to post the letter or you are ordered to burn it', an inference which raises no problems.

University of Nottingham © VERA PEETZ 1979

from the second sentence in (5). Nonetheless, we should not be surprised that the claim made with 'Ralph said that' in (5) is false when the reference of 'that' shifts to (9). The difference is merely that John's utterance of 'Hesperus was visible yesterday' samesays with Ralph's whereas John's utterance of the sentence 'Phosphorus was visible yesterday' does not—even though the second sentence follows from the first using the identity of Hesperus and Phosphorus. However mysterious samesaying has appeared to some, that much should be obvious. In addition, and subject to a caveat which I shall discuss shortly, the following is a 'principle' that holds of the samesaying relation:

> (P) If an utterance x of mine *samesays* with some utterance y of A's and with some utterance z of B's then y samesays with z.

The caveat is this. It might be held that, in spite of the hint of an equivalence relation contained in the 'same' of 'samesays', this relation is not symmetrical. For example, my utterance of 'Something moves' might samesay with your utterance of 'The Earth moves' though circumstances might be such that your utterance does not samesay with mine. That is, it could be the case that your utterance would not be suitable to report the content of an utterance of mine made with the sentence 'Something moves'. Given this sort of case and appealing to the obvious reflexivity of samesaying, a counter-example to (P) could be constructed as follows:

> My utterance of 'Something moves' samesays with your utterance of 'The Earth moves' and my utterance samesays with itself. Hence, by (P), your utterance samesays with mine. But this is precisely what we assumed did not hold. So (P) cannot be generally true if samesaying is ever not symmetric.

However, the following revised version of (P) does seem generally true and, as I shall argue, it does the work I require of it:

> (P') If an utterance x of mine samesays with some utterance y of A's *and conversely*, and with an utterance z of B's *and conversely*, then y samesays with z.

Using (P') I think that the following case leads to trouble for Hornsby's suggestion. There is a man whom I know both as 'Ortcutt' and as 'Fingers'; Ralph knows him only as 'Fingers'. Also, Ralph has swallowed Quine's story and thinks that the man named 'Ortcutt' is a pillar of the community and is very definitely someone other than Fingers. On some appropriate occasion Ralph says: 'Fingers is a crook' and I can report this, truly, using either:

> (i) Ralph said that Fingers is a crook. *Or,*

(ii) Ralph said of Fingers that he is a crook.

Hornsby's account of (ii) would be:

(iii) (∃*u*) (Said (Ralph, *u*, that) & Of (Fingers, *u*)). *x* is a crook.

This comes out true since my utterance of '*x* is a crook' (or, more properly, 'he is a crook'), though open, samesays with Ralph's utterance of 'Fingers is a crook'. Also, Ralph's utterance of 'Fingers is a crook' samesays with my utterance of 'he is a crook'. Both of these samesaying relations are guaranteed by Hornsby's account since both utterances are of Fingers and are the same in import.

Moreover, by this account, it is no less true that my utterance of 'he is a crook' samesays with *my* utterance of 'Ortcutt is a crook', and conversely. But, by (P'), since my utterance of 'he is a crook' samesays *both* with my utterance of 'Ortcutt is a crook' (and conversely) *and* with Ralph's utterance of 'Fingers is a crook' (and conversely), then my utterance of 'Ortcutt is a crook' samesays with Ralph's utterance of 'Fingers is a crook'. This fact should make my utterance of:

(iv) Ralph said that Ortcutt is a crook.

come out true; though it is clearly not.

It may be thought that, in the above, I have been too quick to assume samesaying in both directions and that Hornsby's definition of this relation can be framed so as to make explicit that it holds in one direction only. Some such move might then avoid the above consequence. A moment's thought, however, should convince one of the implausibility of any such move. Her account requires that my utterance of 'he is a crook' samesays with Ralph's utterance of 'Fingers is a crook'; open sentences must be seen, in appropriate circumstances, as samesaying with closed sentences in order to allow for the truth of indirect discourse claims made with the 'saying of' idiom. Further, there can be no acceptable reason for denying that in any such case the relevant closed sentence samesays with the open one. After all, if Ralph had heard me say 'he is a crook' in circumstances where the referent of the pronoun were clearly the man he knew as 'Fingers', then he could truly report 'SG said that I said that Fingers is a crook'. Ralph's complex utterance is true because, in the circumstances described, his closed sentence samesaid with my open one. So, though samesaying may not be symmetrical, there can be no reason for denying that it behaves symmetrically in cases such as this one.

As was noted earlier, there is a reading of (iv) on which it is true but that reading is not the subject of present discussion nor is it one to which Hornsby could appeal to save her account from the above consequence.

In addition, the point of the above example is quite general and cannot be avoided by some simple change in the Hornsby construal of samesaying.

In order to handle the 'saying of' idiom paratactically, samesaying has to be loosened sufficiently to relate open and closed sentences and, however this is done, the embarrassment just described can be made to crop up. A clear view of this can be had by considering why the same problem does not arise for closed sentence samesaying. In the above case I could have truly reported Ralph's utterance using the sentence:

(i) Ralph said that Fingers is a crook.

Given that I know that Fingers is Ortcutt, someone might think that this would entail that my utterance of 'Fingers is a crook' samesays with my utterance of 'Ortcutt is a crook'. If this were so then by using (P') the same embarrassment could be generated; but it isn't so. For, even though I know the truth of the identity in question, the names 'Fingers' and 'Ortcutt' are genuinely different names. That is, they have different senses at least in that they can serve to report other people's sayings either truly or falsely depending on which is used. My knowledge of the identity does not, by itself, suffice to make them notational variants. However, when open sentences are allowed to stand in the samesaying relation to closed sentences, no such appeal to the different senses of names (or different uses) can avoid the abovementioned troublesome consequence.

Nor is the trouble caused simply by samesaying between open and closed sentences. That is, it might be thought that Hornsby's account can be saved by pressing into service Burge's view (mentioned by Hornsby) that proper names are predicates with a demonstrative element and that sentences containing them are open. After all, given Burge's proposals the samesaying relation between sentences containing proper names will be a relation between open sentences only and not between open and closed sentences. However, without speculating on the details of any such use of Burge's views, I think I can show why this will not work in the present case. What Hornsby needs is an account which licenses a samesaying relation between:

(a) 'Fingers is a crook' uttered by Ralph

and

(b) 'he is a crook' uttered by me

but which blocks any such relation between (b) and:

(c) 'Ortcutt is a crook' uttered by me.

If (a) contains an open sentence which, besides the demonstrative element, uses the predicates 'is a Fingers' and 'is a crook', then it cannot

samesay with (b); for (a) may be of the same individual as (b) but it can hardly be the same in import as (b) since the sentence used in (a) contains a predicate not found in (b).

If somehow one disallows such 'name' predicates from upsetting Hornsby's 'same in import' clause, then the relation between (a) and (b) is preserved but at the cost of committing oneself to samesaying between (b) and (c). If 'is an Ortcutt' doesn't count against the same-in-import clause, then (b) is of the same individual as (c) and is the same in import as (c)—and this is just what led to the trouble in the Fingers/Ortcutt case.

Birkbeck College, London

COUNTERFACTUAL ANALYSIS OF CAUSATION AND KIM'S EXAMPLES

By T. YAGISAWA

THE idea of David Lewis's counterfactual analysis of causation[1] is this. For any events c and e, e causally depends on c iff

(i) c and e are distinct events,

(ii) $O(c)\,\square\!\!\rightarrow O(e)$,[2] and

(iii) $-O(c)\,\square\!\!\rightarrow -O(e)$.

c causes e iff c and e are actual events and there is a chain of causal dependence from e to c (i.e., a causal chain from c to e).

It is a consequence of this analysis that for any distinct actual events c and e, if $-O(c)\,\square\!\!\rightarrow -O(e)$, then c is a cause of e. Jaegwon Kim gives apparent examples of such events that are not cause and effect.[3]

1 If I had not written 'rr', I would not have written 'Larry'.

2 If yesterday had not been Monday, today would not be Tuesday.

3 If George had not been born in 1950, he would not have become 21 in 1971.

4 If John's sister Ann had not given birth at t_1, he would not have become an uncle at t_1.

5 If Socrates had not died at t_2, Xanthippe would not have become a widow at t_2.

6 If Bill had not raised his arm at t_3, he would not have signalled at t_3.

---

[1] 'Causation', *Journal of Philosophy*, LXX, 17 (Oct. 11, 1973), 556–567.

[2] The notation is Lewis's: $O(e)$ is the proposition that e occurs; if p and q are propositions, '$p\,\square\!\!\rightarrow q$' expresses the counterfactual conditional that if p were true, then q would be true.

[3] 'Causes and Counterfactuals', *Journal of Philosophy*, LXX, 17 (Oct. 11, 1973), 570–572.

These examples are not exactly the same as the ones actually given by Kim, but they fairly represent the spirit.

In this paper I will suggest one way to handle these counterexamples. I will assume that all the events involved in the examples are actual. I shall first propose to revise the original counterfactual analysis of causation slightly and proceed to show that the counterfactual dependences of 1–6 are not causal according to the revised analysis. Several comments will then follow.

The Revision: Reformulate the definition of causal dependence as follows. For any event c and e, e causally depends on c iff

(i)′ $-O(c)$ and $O(e)$ are compossible, i.e., there is a possible world in which $-O(c)$ and $O(e)$ are both true

(ii) $O(c) \square \rightarrow O(e)$, and

(iii) $-O(c) \square \rightarrow -O(e)$.

The rest of the analysis remains the same as the original. (Note that distinctness of c and e is guaranteed by (i)′.)

Thus revised, the analysis still gives a necessary condition for causation. For example, I threw a ball and it broke the window: the very same particular breaking of the window could have happened without the very same particular throwing of the ball by me; a different throw of a ball by me or someone else, or a throw of a stone by me or someone else, or something entirely different from any throwing such as a shock wave, could have caused the same breaking of the window.

It should be obvious that the revised analysis excludes examples 1 and 2. My writing 'rr' was actually a part of my writing 'Larry' so that the second could not have occurred without the first occurring: any writing of 'Larry' that does not include writing 'rr', if such a writing is really a writing of 'Larry' at all, would not have been the same as the actual writing of 'Larry' by me that actually included writing 'rr'. Yesterday's not having been Monday is not compossible with today's being Tuesday. Example 3 is an odd case. If it is about actual particular events of George's birth and of his becoming 21, then the original analysis by Lewis will exclude it: it will fail to satisfy (iii) of the original analysis, for if that particular birth of George had not occurred and some slightly different birth of George had occurred at the same time as the actual birth, then still there would have occurred the same particular becoming 21 of George. If 3 is not about that actual event and that becoming, what is it about? Probably it is merely about George's having been born in 1950 and his having become 21 in 1971. If so, then the spirit of the revised analysis will exclude it; George's not having been born in 1950 is not compossible with his having become 21 in 1971.

Consider   5a Socrates did not die at t_2.
                  5b Xanthippe became a widow at t_2.

5a and 5b are compossible, for Xanthippe could have had a different husband, who died at t_2. But it is no objection to the revised analysis that they are. A real objection would have to show 5a' and 5b' to be compossible:

5a' Socrates' death at t_2 did not occur,

where 'Socrates's death at t_2' rigidly refers to the actual death of Socrates: so 5a' is different from 5a''

5a'' *A* death of Socrates did not occur at t_2.

5b' Xanthippe's becoming a widow at t_2 occurred,

where 'Xanthippe's becoming a widow at t_2' rigidly refers to the actual becoming a widow of Xanthippe: so 5b' is different from 5b''

5b'' *A* becoming a widow of Xanthippe occurred at t_2.

What is at stake is causal *relations* among *particular* events, not causal regularities among event *types*. Let me make this explicit. Let 'AB' be a name for the actual particular event of Ann's giving birth at t_1, 'JU' for the actual particular event of John's becoming an uncle at t_1, 'SD' for the actual particular event of Socrates's death at t_2, 'XW' for the actual particular event of Xanthippe's becoming a widow at t_2, 'BR' for the actual particular event of Bill's raising his arm at t_3, and 'BS' for the actual particular event of Bill's signalling at t_3. The question is whether propositions $-O(AB)$ and $O(JU)$ are compossible, whether $-O(SD)$ and $O(XW)$ are compossible, and whether $-O(BR)$ and $O(BS)$ are compossible.

Thus far I have been talking as if becoming an uncle, becoming a widow, and signalling, are events. I now want to set myself free from this presupposition and allow a possibility of their being something different from events. However, I do not intend thereby automatically to be able to cope with examples 4–6. Accordingly, I shall modify the revised analysis in such a way that 'c' and 'e' now range over not only events but also entities like becoming an uncle, becoming a widow, signalling, and so on. (It is a general assumption of the present paper that they, as well as events, are entities of some sort or other over which we can quantify and that they may stand in a causal relation to something.) Also, I'd like to use the locution 'happen' for such entities as JU, XW, and BS to distinguish them from events, entities for which we have the locution 'occur'. So in the revised analysis '$O(x)$' should be reinterpreted as 'x occurs or happens'.

I think that apparent compossibilities of those pairs of propositions

above are due to the obscurity of our conception of what JU, XW, and BS really are. They are particulars all right, but of what kind? In my opinion, the lesson we should learn from Kim's examples is that we should ask for a clear conception of such particulars. I claim that there is an adequate theory of particulars like JU, XW, and BS, such that under that theory the above three pairs of propositions are not compossible. For the sake of simplicity I shall talk as if identical objects, events, . . . do exist, occur, . . . in different possible worlds and ignore counterpart theory. The essence of what I will say does not depend on this simplification. I assume that names 'AB', 'JU', 'SD', 'XW', 'BR' and 'BS', are all rigid.

Let us first note that JU, XW, and BS, are similarly peculiar kinds of particulars. Each of them somehow involves a certain event and a certain state of affairs. JU involves AB and the state of affairs, let us call it S_4, of Ann's being John's sister. XW involves SD and the state of affairs, called S_5, of Socrates's being Xanthippe's husband. BS involves BR and the state of affairs, S_6, of a certain convention of signalling being agreed upon. I call JU, XW, and BS, *complexes*. Complexes are such a kind of entity that their ontological mode is happening, which is characterized by the following principles.

> For any complex c, there are a certain event e and a certain state of affairs s such that for any time t c happens at t iff e occurs at t and s holds at t.

A complex happens iff it happens at some time.

(I have ignored non-essential complications involved in the case of a complex happening during a time interval.) If c is a complex which happens at t iff event e occurs at t and state of affairs s holds at t, it is convenient to identify c with the pair $\{e,s\}$. Such an identification is only a handy way of referring to complexes, and nothing more: it should not be taken to commit one to the view that complexes are sets. Complexes can be any kind of entity you wish provided that they meet the above principles. I identify JU, XW, and BS, as follows.

$JU = \{AB, S_4\}$.
$XW = \{SD, S_5\}$.
$BS = \{BR, S_6\}$.

It is now obvious that for any possible world w, if AB did not occur in w, then JU did not happen in w. Thus, according to the revised analysis, adequately modified as indicated earlier, AB did not cause JU. Similarly with the other cases.

We also have the following results.

4.1. Suppose in a possible world w Ann did not give birth but Rebecca,

another sister of John's, did. Then, JU did not happen in w: because JU happened in w only if AB occurred in w but AB did not occur in w. Notice that in w it is still true that John became an uncle, for in w *some* becoming an uncle of John happened, i.e., it is true in w that $(\exists u)$ $(\exists x)(\exists y)(\exists z)$ (z is John's sister & y is the state of affairs that z is John's sister & x is a giving birth by z & $u = \{x,y\}$ & u happened). Generally, a person j becomes an uncle iff $(\exists u)(\exists x)(\exists y)(\exists z)$ (z is j's sister or sister in law & y is the state of affairs that z is j's sister or sister in law & x is a giving birth by z & $u = \{x,y\}$ & u happens).

4.2. Suppose in a possible world w Ann did not give birth at t_1 but she did at some other time. Then, whether JU happened in w depends on whether Ann's giving birth in w is the same particular event as AB, the actual giving birth by Ann.

5.1. If in a possible world w Xanthippe had a different husband, who died at t_2, then XW did not happen at t_2 in w, for S_5 did not hold at t_2 in w. Nevertheless it is true in w that Xanthippe became a widow at t_2, for *some* becoming a widow of Xanthippe happened at t_2 in w. Generally, a person j becomes a widow at t iff $(\exists u)(\exists x)(\exists y)(\exists z)$ (z is j's husband at t & y is the state of affairs that z is j's husband & x is a death of z & $u = \{x, y\}$ & u happens at t).

5.2. If in a possible world w Socrates was Xanthippe's husband until he died and he did not die at t_2 but died at t_3, then whether XW happened at t_3 in w depends on whether that death of Socrates in w is the same particular event as SD.

6.1. If in a possible world w Bill did not raise his arm but did something else at t_3 that counted as signalling by some agreed convention, then BS did not happen at t_3 in w, for BR did not occur at t_3 in w. Nevertheless it is true in w that Bill signalled at t_3, for *some* signalling by Bill happened at t_3 in w. The general condition for signalling is obvious.

My proposal hinges on an intuition that JU cannot happen without AB occurring, and similarly in the other cases. This is an intuition I have, and seems to me to be quite a natural one. Although to those who do not share my intuition I can offer no decisive argument to convince them into it, I can say this much: since there is a strong semantic connection of synonymy between 'becoming an uncle' and 'sister or sister in law's giving birth', between 'becoming a widow' and 'husband's dying', and between 'signalling' and 'doing the right kind of thing according to a certain convention of signalling', it should not be surprising that there is a strong ontological connection between a particular becoming of an uncle and a particular giving birth, between a particular becoming of a widow and a particular death, and between a particular signalling and a particular action.

The difference between complexes and events should be clear by now. In a sentence a complex ontologically depends on a certain particular

NOTES

The ANALYSIS Committee consists of: Chairman, P. T. Geach; Secretary, J. H. Benson; T. R. Baldwin, A. E. Fisher, Andrew Harrison, R. F. Holland, Hidé Ishiguro, J. Kemp, J. H. McDowell, Bernard Mayo, D. H. Mellor, R. G. Swinburne, A. R. White, C. J. F. Williams, Peter Winch. This committee is responsible for appointing and advising the Editor and for the general policy of the paper.

SUBSCRIPTIONS. The subscription to ANALYSIS for institutions is £5.00 (inland), £6.00 (overseas), $15.00 (U.S.A. and Canada); for individuals £4.00 (inland), £4.80 (overseas), $12.00 (U.S.A. and Canada). Each volume comprises four numbers, three of 48 pages and one of 64 pages, appearing within the academic year—in October, January, March and June. Orders should be sent to Basil Blackwell, 108 Cowley Road, Oxford OX4 1JF, or placed with any bookseller.

CONTRIBUTIONS. **Articles submitted for publication should be addressed to Mr. Christopher Kirwan, Exeter College, Oxford OX1 3DP.** Contributors are asked to note the following requirements.

Articles should **normally** not exceed 3,000 words in length. **Occasionally,** however, longer contributions can be accepted;

They must be type written in double spacing on one side of the paper only;

Footnotes should be kept to a minimum and wherever possible avoided altogether;

Single quotation marks should normally be used, except for purposes of internal quotation and "scare" quotes.

Discussion papers should be sent in as soon as possible after the appearance of the article to which they refer.

It is regretted that owing to increased postal charges it is no longer possible to return typescripts unless the following instructions are followed:

Contributors in the United Kingdom should enclose a stamped addressed envelope of suitable size; if immediate acknowledgement is required, a stamped postcard should also be enclosed.

Overseas contributors who wish to have their MSS. returned should send an envelope **and international reply coupons** of the requisite value, whether for air or surface mail.

Galley proofs of accepted articles will be sent to authors for correction, together with information about offprints. Typescripts will be retained by the Editor on the assumption that authors have kept their own copies.

The copyright of articles printed in ANALYSIS remains the property of the author, but contributors are strongly advised, in their own interest, to consult the Editor before consenting to the reprinting of their articles.

ANALYSIS

Edited by

CHRISTOPHER KIRWAN

CONTENTS

REPORT ON *ANALYSIS* "PROBLEM" No. 17

By G. E. HUGHES

> Can I ever, by my subsequent actions, bring it about that something I did
> on a previous occasion was done from a certain motive rather than from
> some other one?

MY method of selecting a competition problem was first to make a
list of a number of problems that interested and puzzled me, then
to toss them as topics of conversation to a number of colleagues (to
whose patience and forbearance I hereby publicly pay tribute), and
finally to choose the one that seemed to provoke the most interesting
discussions. Nevertheless the Editor informs me that there were no
entries at all. Possibly what excites and intrigues philosophers in
Wellington bores or baffles philosophers elsewhere. At any rate, I feel
I can hardly commend my method to future competition-setters.

I shall try to indicate briefly what I had in mind in composing the
question and why it seems to me important.

One thing that seemed to me (obviously incorrectly) to be a good
augury for my topic was that the conversations I had about it tended to
begin with an 'Of course not' and end with a 'Probably yes'. The initial
impulse to say 'Of course not' stems, I think, from the idea that an
affirmative answer would involve a claim to be able to change the past.
Now the question 'Can we change the past?' seems to me to be less
open-and-shut than it is commonly taken to be, but in fact the problem
was carefully worded so that it did *not* raise it. At least it did not raise
it if by 'changing the past' we mean something like bringing it about
that some event that really did occur did not occur after all, or that it
happened differently from the way in which it really did happen.
Whether the notion thus vaguely delineated can be given a clear and
consistent sense is a nice question. My problem, however, did not
involve any such notion, but only that of doing something that gives to
a past event some characteristic that it would not otherwise have had.

Now for some characteristics at least, this notion is not a problematic
one at all. It is often pointed out, indeed, that we are all doing this kind
of thing all the time and on a vast scale. Last year, for instance, I brought
it about that Julius Caesar's first visit to Britain had the characteristic
of taking place 2032 years before my first visit to Kathmandu. Of
course people want to protest that that isn't a real characteristic of
Caesar's visit to Britain, so let us consider examples that may appear
intuitively more substantial (merely remarking in passing that it is
desperately difficult to say just what would and would not count as a

real characteristic in the sense that the protesters seem to have in mind). Mid-way through delivering a lecture it may occur to me that if I continue in one way my opening remarks will have formed a good introduction to the lecture and that if I continue in another way they will not. In that case I can bring it about that those opening remarks, which lie in the past, were a good introduction or that they were not. Again, it seems clear that a remark made in conversation can sometimes be insulting if the speaker says nothing further but not insulting if he follows it up in a certain way. So if I have made such a remark I can at least sometimes bring it about, by what I subsequently say or do not say, that it was, or that it was not, an insult.

Can we say something analogous about motives? One might argue that we can along these lines. It seems altogether unsatisfactory to say that what motive a person acts from is always determined solely and unambiguously by the thoughts that were in his mind at the time of acting, or even by the events in his life-history that led up to the action. We ascribe motives to people not solely on the basis of our beliefs about such things but also on the basis of our beliefs about what they would do if certain situations were to arise; and at least sometimes, if these beliefs about what they would do if . . . turn out to be false, we regard this as showing that our original ascription of the motive in question was incorrect. One may find oneself wondering whether someone performed a certain action out of spite, or envy, or perhaps even out of affection, or a sense of duty; and it may be that the whole of his occurrent life-history to date is consistent with each of these and indeterminate as between them, so that the only thing that would settle the matter is what he would do if . . . But now this may be something that is within his control; for he may find himself in, or get himself into, situations of the appropriate kind and then *decide* what he will do.

Or would that not count? I can imagine someone arguing that anything he does later on can tell us at most something about what the pattern of his motivation has by then become and not what the motive of the original action was, that what is relevant to the latter is not what he would do if . . . but only what he would *have done* if . . ., and that his subsequent actions have no bearing on that.

I do not know how to assess these opposing arguments with any confidence. I suspect, too, that my question is one to which neither a direct affirmative nor a direct negative answer is simply correct, but that one answer may nevertheless be better than the other. At present I am inclined to think that 'Yes' is a better answer than 'No'; but I still have qualms, and I had hoped that some of the entries would help me to get clearer about the issue.

I think the question is important for this reason. There is widespread agreement that some motives are morally better than others. So if we

give an affirmative answer to my question it looks as if we should recognise a class of duties which are not, I think, usually mentioned in the text-books, *viz*. duties to bring it about that we *shall have* acted from better motives rather than from worse ones.

Victoria University of Wellington

SINGER'S CRITIQUE OF THE MARKET

By Tom Regan

IN his 'Rights and the Market'[1] Peter Singer launches a series of objections against those he calls 'defenders of the market' (p. 215), persons who oppose government intervention in the free exchange of goods and services. One of Singer's criticisms in particular deserves attention since, if it happened to be sound, it would show that defenders of the market are confused about the very thing they claim to prize most highly—namely, individual liberty.

Singer's criticism takes the following form. Defenders of the market, he writes (ibid.), 'regard every law extending the range of choices formally open to people as an increase in their freedom, and every law diminishing this range of choice as a decrease in freedom.' Serious questions must arise concerning the accuracy of Singer's attributing this view to all defenders of the market; for example, a law which increases the range of choices formally open to *some* people may violate the rights of others and thus would not be sanctioned by a defender of the market such as Nozick.[2] But setting these matters to one side, Singer disputes the view just characterized by arguing that situations can arise where a law can 'in one sense' diminish our range of choices without at the same time decreasing our freedom. The general type of case Singer has in mind is one in which the choices of an individual, considered in isolation from the choices other individuals might make, may be rational, but, when considered in conjunction with the choices of others, 'the cumulative effects may be disastrous for everyone' (ibid.) An individual's use of his/her private car may be like this. Considered in isolation, John's and Jane's use of their cars seems quite harmless; when considered in conjunction with the transportation choices of others, however, long

[1] In *Justice and Economic Distribution*. John Arthur and William H. Shaw (eds.) Prentice Hall, Inc. (Englewood Cliffs, New Jersey) 1978, pp. 207–221. Page references henceforth are included in the body of my essay.
[2] *Anarchy, State, and Utopia*. Basic Books (New York) 1974.

, delays, serious problems of pollution, etc. might result. In such a case, Singer contends, legislation prohibiting the use of private vehicles in a given urban area, together with the development of a cheap, reliable and speedy mass transit system, might be in everyone's interest. 'In one sense,' Singer writes (p. 216), 'the range of choice of transport has been reduced'—that is, a choice previously left to the individual's discretion, free of coercion (to drive or not to drive in the specified urban area), has been eliminated—'but on the other hand,' he continues, 'a new choice now opens up to us—the choice of using a fast and frequent public transport system at a moderate cost.' Thus, here we have a law which, though it diminishes our 'range of choice' 'in one sense,' does not represent a 'decrease in freedom' all considered. A previously existing choice may be lost, but a new one is gained.

The general point of Singer's criticism, then, seems to be this. Defenders of the market fail to see that legislation can decrease the range of our choices without thereby decreasing our freedom. Accordingly, the market's defenders are confused when they oppose coercive legislation on the grounds that it decreases individual freedom. Such legislation can create 'new choices' and thus foster rather than decrease our liberty.

This criticism of Singer's seems to rest on a failure on his part to distinguish clearly between (a) having a choice and (b) having the freedom to choose. Consider the case of the mugger and the muggee. Following Singer, one could say that, though 'in one sense' the range of choice open to the muggee has been decreased by the presence in his face of the mugger's pistol, 'on the other hand a new choice now opens up to him (i.e., the muggee)'—namely, to hand over his wallet or resist. Thus, again following Singer, we could say that, though the muggee has lost a previous choice, a new choice has been gained, and the scales of freedom remain in equipoise.

Now, it is clear that the mugger opens the eyes of the muggee to a 'new choice'. But it is equally clear that the muggee is not thereby the beneficiary of a new freedom, let alone one that nicely cancels out the lost freedom symbolized by the pistol in his face. In general, people placed in circumstances where they are coerced to choose between alternatives none of which they would choose if left to their own desires or preferences, where previously coercion was absent, *do* have 'a new choice'—namely, to choose one or the other of the unwanted alternatives. But the coming into being of this 'new choice' is not the birth of a 'new freedom'. People are not at liberty to choose (do not have the freedom to choose) when they are made to choose what they do not want. To think otherwise is to suppose that freedom of choice increases proportionately to the amount of coercion, which is absurd.

·   ·· Possibly it will be objected that the mugger-muggee example beclouds the issue. After all, the muggee hardly stands to gain from his

transaction with the mugger, whereas, according to Singer, each owner of a private vehicle can view legislation restricting its use as in his/her interest.

At least three replies come to mind. First, a lot of issues go begging in Singer's own example. Some people *would* rather sit in traffic jams, etc. instead of using quick, cheap and efficient public transportation. *Some* people would *rather* travel alone. To tell them that restrictive legislation making the private use of their car illegal in certain areas 'does not diminish their freedom' and 'is in their self interest' is not likely to set well. Such talk has the ring of the language of the bureaucrat who masks the abrogation of individual freedom by calling it by another name. Like the muggee, the man who *wants* to use his car even if it means long delays, etc. gets a 'new choice,' when the law takes effect, but he does not thereby get a 'new freedom' in the bargain.

Second, even if it were true that the growth of coercive legislation benefited everyone (had what Singer might term 'attractive' consequences, p. 215), that would not guarantee that individual liberty was not a casualty. Certainly it is possible that a mass transit system, to stay with Singer's example, might in some sense be in everyone's self-interest, might in some sense make everyone's life happier. Nevertheless, it does not follow that it would make anyone's life freer. Whether it does or not is an open question, one that is not answered at all by noting that, with the arrival of the buses or whatever, everyone will have 'new choices.'

But third, even if the consequences of a piece of coercive legislation were 'attractive' for everyone and, indeed, actually did bring about an end-state where individuals enjoyed an increase in personal freedom, that would not satisfy Nozick for one, or at least Nozick as he is interpreted by Singer. (And it is Nozick who, as a 'defender of the market,' is Singer's principal target). Nozick, according to Singer (p. 208), 'rejects altogether the idea that institutions . . . are ultimately to be judged by the ends they promote,' including the end of 'the maximization of freedom'. So, *either* Singer's speculation about the mass transit example (and similar cases) fails to illuminate how a decrease in the range of our choices does not represent a decrease in our freedom (the general point of my first two replies) *or*, conceding that it does, it fails seriously to engage in debate with the particular 'defender of the market' at whom the argument is principally aimed.

North Carolina State University   © TOM REGAN 1979

REGAN'S CRITIQUE OF SINGER

By PETER SINGER

'POSSIBLY it will be objected' says Tom Regan in his critique of my discussions of freedom and the market, 'that the mugger-muggee example beclouds the issue.'

Well, that thought did cross my mind. In setting up the example of a choice between an hour's crawl along congested, exhaust-filled roads and 20 minutes comfortable bus ride, I supposed that 'most reasonable people . . . would have little hesitation in choosing the latter'. For this reason, I argued, if we suppose that an efficient bus service can survive only when the use of private vehicles is restricted, a law restricting the use of private vehicles would not diminish freedom, since the possibility it restricts (driving one's own car) is balanced, if not outweighed, by the possibility it creates (going on a quick and cheap bus).

How is the example of mugging supposed to be analogous to this? The mugger obviously eliminates a possibility most people would choose if they could, namely retaining both wallet and life. Those who do not want their wallets can give them to passing strangers at any time, and those who want to die can usually arrange this too. I am not clear what compensating new possibility Regan thinks the mugger provides, but it does not seem to be one that many would choose voluntarily.

Wisely foreseeing opposition to his counter-example, Regan makes three replies:

(1) Some would rather travel alone, however slowly, than by bus.

There may be some people with this preference. If so, I would not deny that their freedom has been diminished by the restriction on the use of private cars. But if, as I supposed in the example, most people have the opposite preference, and if it is impossible to satisfy both preferences simultaneously, then my claim that freedom is, *on balance*, not diminished by the restriction still stands.

(One word of apology: what I actually wrote on p. 216 of 'Rights and the Market' was that 'most reasonable people' would choose the bus. I should simply have said 'most people', for the use of 'reasonable' may have misled Regan into thinking that I intended to disregard the preferences of those who enjoy traffic jams on the grounds that anyone with such peculiar preferences must be unreasonable. This move would indeed have 'the ring of the language of the bureaucrat who masks the abrogation of individual freedom by calling it by another name'; but it was not a move I ever intended to make. Of course, that most people prefer fast buses to slow cars is merely a supposition for the sake of the example —as with most philosophical examples, its accuracy is irrelevant to the

general philosophical point the example was designed to make.)

(2) Even if coercive legislation benefited everyone, liberty might still be a casualty.

I agree. Some paternalistic legislation—like making it compulsory to wear seatbelts—benefits people by compelling them to do something they would not do in the absence of legislation. My argument did not invoke paternalistic legislation, however, but legislation which makes it possible for people to do what they wish to do. My claim that such legislation can increase freedom has nothing to do with the quite separate issue of whether such legislation benefits people. In practice, people usually wish to do what they can see to be in their interests, and the fact that something is in people's interests is a good—but not conclusive— reason for supposing that once they see this, they will choose it. In my example, I supposed that people would choose what I took to be in their interest, and it was on their supposed choice, rather than their supposed interest, that the argument rested.

(3) My argument does not work against its principal target— Robert Nozick—since, on my own showing, Nozick is concerned with the non-violation of rights rather than the maximization of freedom.

I have two rejoinders to this. First, though Nozick may have been the principal target of my article, he was not the only one, and the particular argument to which Regan objects was aimed, as its footnote indicates, at a view of liberty held by F. A. Hayek and Isaiah Berlin. Second, as the same footnote goes on to point out, Nozick holds that rights are not infringed by the collective effect of several individual actions, though if a single action had the same effect, it would be an infringement of rights. This claim, which is crucial to Nozick's defence of the market against the standard socialist objection that it deprives the hapless individual worker of the right to earn a living wage, will seem less plausible if we come to see social arrangements, and not just deliberate human actions, as forces restricting freedom. Nozick's view of rights is structurally similar to Hayek's view of freedom.

So neither Regan's counter-example, nor his supporting objections, shake my argument that legislation may increase freedom by restricting choice. The essential point is simply that it is not only laws and deliberate human acts which reduce our freedom, but also circumstances, both natural and social. If I did not make this point clearly enough in the original article, I am thankful to Regan for forcing me to restate it.

Monash University   © PETER SINGER 1979

THE FIRST PERSON PRONOUN:
A REPLY TO ANSCOMBE AND CLARKE

By MICHAEL J. WHITE

IN a 1975 paper, 'The First Person',[1] Professor G. E. M. Anscombe concludes that 'I' is not a referring expression. While Anscombe admits that many I-propositions (e.g., that expressed by a use of the sentence 'I am sitting') can be verified or falsified by the examination of a certain object or "thing" (e.g., an animate body), such a proposition expresses—evidently not in a subject-predicate form—an 'unmediated agent-or-patient [conception] of actions, happenings, and states' (p. 65). D. S. Clarke, Jr. has, in a recent issue of ANALYSIS,[2] defended Anscombe's conclusion concerning the non-referential status of 'I' but has argued, *contra* Anscombe, that proper nouns and demonstrative phrases can be correctly substituted for the first person pronoun in certain contexts.

Anscombe's argument (and, I suspect, Clarke's defence of it) seems motivated by the desire to avoid the necessity of a search for a metaphysical (and perhaps mysterious) "self" to serve as the referent of a use of 'I'. If 'I' does not refer at all, we are obviously relieved of the need to find any such special referent. I am in sympathy with Anscombe and Clarke's apparent desire to avoid being "forced" by a referential semantic conception of 'I' to some specific conclusion concerning the "real self"; furthermore, the connection Anscombe draws between 'I' and 'unmediated agent-or-patient conceptions of actions, happenings, and states' strikes me as fundamentally correct. None the less, the denial of a referent to (a use of) 'I' seems unintuitive, particularly in view of the fact that neither Anscombe nor Clarke really gives an alternative account of the logical form of sentences containing the first person pronoun. It also seems unnecessary. I shall argue (1) that it is not necessary to give up the referring conception of 'I' in order to avoid the imposition of a particular metaphysical notion of the "self" and (2) that a referring conception of 'I' can nicely accommodate Anscombe's notion of the connection between 'I' and 'unmediated agent-or-patient conceptions of actions, etc.'

Before turning to what I take to be Anscombe's principal argument, I should perhaps address a preliminary but fundamental issue. It might be thought—though I do not attribute this view to either Anscombe or Clarke—that 'I' is not a referring expression simply because it seems

---

difficult or impossible to find a "natural language" phrase that obviously does refer and that is synonymous with 'I'. 'I', for example, is not synonymous with 'M. J. White'. Why not? Because 'I' is (if it refers) indexical; if it has a referent at all, the referent that it has depends upon who is "using" it. Its referent is dependent on context in a way that the referent of 'M. J. White' is not. Similarly I agree with Clarke's claim (p. 93) that 'I' is not synonymous with 'this speaker' or 'the person producing this utterance'.[3] *In certain contexts* 'I' (if it refers at all) and one of the latter two phrases will refer to different entities. However, it is still possible to argue as follows: 'I' *does* refer relative to a context. Furthermore, it refers relative to a context in a way different from the way certain other indexical demonstrative phrases (e.g., 'this speaker') refer relative to context. Yet, in a certain context, 'M. J. White', 'I', and 'the person writing this sentence' may all share the same referent.

Anscombe, I believe, would admit that if 'I' were merely an indexical, even an indexical for which no synonymous phrase that *is* a referring expression could be found, that fact would not entail that 'I' does not refer. In fact, her most persuasive argument for the non-denoting character of 'I' involves the attempt to construct a scenario showing that the claim that 'I' is not a referring expression amounts to *more* than the trivial claim 'that we perhaps would not *call* a word a proper name if everyone has it and used it only to speak of himself' (p. 48).

The scenario posits a society in which everyone is labelled with two names. One type (what I will call the β-type, including names '*B*' through '*Z*') is placed on the bearer in such a way that he cannot see it but others can. A β-name is *proper* to its bearer, varying from person to person. The other, α-type, name is stamped on the inside of the bearer's wrist where he can refer to it at will and is the same ('*A*') for everyone. What is the difference between '*A*', which is *ex hypothesi* referential, and 'I', which Anscombe argues is not referential. Anscombe locates the difference in terms of two senses of 'guaranteed reference'.

The first sense of the phrase entails merely that there is an *existent* object 'meant by the speaker' whenever a term X is used by someone

---

[3] Although I agree with Mr. Clarke's conclusion, I am not persuaded by his argument that 'I' is not synonymous with 'this speaker' *because* 'I am speaking' is contingent but 'This speaker is speaking' is analytic. Particularly when dealing with indexicals, I prefer to follow the usage of David Kaplan ('The Logic of Demonstratives', Unpublished Mimeo, 1971, expanded version, 1973), who regards a *sentence* (more precisely a *sentence character* or "meaning") as analytic if and only if it expresses some true proposition or other in every possible context of use, and a *proposition* as necessary if and only if it is true in all possible worlds (accessible to the one from which it is being evaluated). According to this usage, 'I am speaking' would be analytic (would express a true proposition in every possible context of employment), but none of these propositions would be necessary. Given an indexical, *referential* reading of 'this speaker', we might conclude that 'This speaker is speaking' is similarly analytic but does not, in any given context of use, express a necessary proposition. An *attributive* reading of 'this speaker' (e.g., one that ultimately reduces the sense of 'This speaker is speaking' to something like 'For all x, if x is the object of my demonstration and x is speaking, then x is speaking') may result in the claim that 'This speaker is speaking' expresses, regardless of context, a single necessary proposition.

who grasps the correct use of X. 'I', if it refers, has this sort of guaran-
teed reference. However, as Anscombe admits, so does '*A*': 'the "*A*"-
user means to speak of a certain human being, one who falls under his
observation in a rather special way. That person is himself, and so,
given that he has grasped the use of "*A*", he cannot but be speaking
of a real person' (p. 57).

Anscombe distinguishes, however, a stronger sense of 'guaranteed
reference' in which the correct use of a term X guarantees 'not just that
there is such a thing as X, but also that what I take to be X *is* X' (ibid.).
The form of Anscombe's consequent argument is that of a *reductio*.
'*A*' does not possess this stronger sort of guaranteed reference. However
if 'I' is a referring expression at all, it does possess this sort of guaranteed
reference. But the possession of guaranteed reference of the strong
variety requires such stringent identity conditions for the supposed
referent of 'I' that we are led to Descartes' conclusion about the nature
of the referent along with all of that conception's attendant difficulties
(e.g., the problem of the reidentification of the "self" in different
I-thoughts). But the Cartesian conception of the "self" is false (or
incoherent). Therefore, Anscombe concludes, 'I' is not a referring
expression at all.

I wish to block this *reductio* argument by claiming that 'I' no more
has guaranteed reference in the strong sense (if it refers) than does '*A*'.
Situations in which an *A*-user might be mistaken about the referent of
one of his uses of '*A*' possess analogues in which an I-user might be
mistaken about the referent of one of his uses of 'I'.

B (an *A*-user) is shown a colour photograph of a wrist that very
much resembles his own. B's wrist was photographed yesterday and he
believes (falsely) that he is being shown a photograph of B's (his own)
wrist. B comments, '*A* evidently—though *A* did not notice it then—
had a little sunburn yesterday'. However, what B says is false: the wrist
belongs not to B (a 'certain human being, one who falls under [B's]
observation in a rather special way' [p. 57]) but to C.

An analogous case is easily constructed for 'I'. White (an I-user) is
listening to a voice on a tape recording reading a lecture he (White)
taped yesterday. Unbeknownst to White, a talented mimic has also re-
corded his lecture and White is listening to *that* lecture. White comments,
'I evidently—though I did not notice it then—was hoarse yesterday'.
This, in fact, is false.

Wherein lies the error in each case? I think the most plausible answer
is that the *A*-user (B) and the I-user (White) make analogous mistakes.
The *A*-user holds that the sentence

  (1) The person in the picture had a sunburn yesterday

expresses a true proposition. He also holds that the sentence

(2) The person in the picture is (identical with) A (i.e., the individual 'who falls under [the A-user's] observation in a rather special way')

expresses a true identity proposition. Using the principle of 'Indiscernibility of Identicals' the A-user infers that

(3) A had a sunburn yesterday.

Similarly, the I-user holds that the sentence

(1') The person speaking in the recording was hoarse yesterday.

expresses a true proposition. A true identity proposition is also expressed, he believes, by

(2') The person speaking in the recording is (identical with) I (i.e., the individual [oneself] that the I-user stands in a unique, 'unmediated' epistemic relation to).

He infers that

(3') I was hoarse yesterday.

I conclude that Professor Anscombe has not successfully established a significant difference between the A-user's use of the indexical referring expression 'A' and the I-user's use of the first person singular pronoun. However, Anscombe seems to me to be correct in claiming that I-use is somehow connected with 'unmediated agent-or-patient conceptions of actions, happenings, and states'. I suggest that this insight can be reconciled with the referential interpretation of 'I'. The referent of 'I', as used by individual X, is the "ultimate subject" of X's actions, passions, etc. This characterization leaves open the "nature" or "identity conditions", if you will, of the proper referents of first person pronouns. A metaphysical question, if indeed this be one, is *not* thereby settled by a referential semantic account of the first person pronoun. Fortunately, we can perfectly well get on, employing the first person pronoun quite correctly, without supplying any such identity conditions for "selves". The fact that we do not perhaps have, in this metaphysical sense of 'self', a 'distinctly conceived' (p. 65) referent for a use of 'I' does not entail that the pronoun, as employed by a given person (or even a given machine),[4] has no referent.

[4] It seems to me that the issue of "self-consciousness" is psychological or, perhaps, metaphysical in nature. This issue too, in other words, need not be decided in a particular way by an adequate referential semantic account of 'I'. According to the referential semantic account of 'I' alluded to in this paper, there would be no *special* problems in programming a machine to use 'I' correctly, i.e., to use it with the same *semantic* force as that with which English speakers use it.

Arizona State University   MICHAEL J. WHITE 1979

CURRY'S PARADOX

By ROBERT K. MEYER, RICHARD ROUTLEY *and* J. MICHAEL DUNN

A PROJECT long dear to the hearts of some logicians has been the formulation of a naive set theory in which the full comprehension principle holds—i.e., in which for every open sentence $A(x)$ in one variable x there exists the class $\{x : A(x)\}$ consisting of exactly the things that satisfy $A(x)$.

Russell's paradox put a well-known crimp into this project. Let us review the argument. Using '$\longleftrightarrow$' for logical equivalence, a naive set theory surely has the

Abstraction principle. $\quad t \in \{x : Fx\} \longleftrightarrow Ft$

for all terms t and open sentences Fx. As an instance we have

Russell's paradox. $\{x : x \notin x\} \in \{x : x \notin x\} \longleftrightarrow \{x : x \notin x\} \notin \{x : x \notin x\}$;

i.e., using 'R' for '$\{x : x \notin x\}$'

(1) $R \in R \longleftrightarrow R \notin R$

(1) looks pretty bad, and indeed it has been sufficient to give a truly naive set theory a bad name. The usual conclusion drawn, as the reader is aware, is that not every open sentence determines a class t satisfying the abstraction principle.

That conclusion, however, has never been wholly satisfactory. The conspicuous virtue of talk of classes is that it provides us, at least imaginatively, with reified predicates, and it has always been disturbing to hear that Russell's paradox "proves that" some predicates cannot be so reified. Set theory is, at best, an exercise in philosophical and mathematical imagination anyway, and any impediment to the scope of that imagination must sometimes appear *ad hoc* and artificial. (Thus, e.g., for all the talk of "constructing" sets according to some formal recipe or other, the constructing takes place at best in our heads, only unwashed Platonists daring to hold that anything in reality answers to what is cooked up by such recipes. While it is interesting that some of these recipes—e.g., those of Zermelo-Frankel, and of type theory—are relatively faithful to intuitive ideas exploited in their motivation, while others—e.g., Quine's *New Foundations*—are just odd, we are nevertheless impressed with Quine's philosophical arguments that, though our intuitions are "bankrupt" here, that set theory is best which meddles least with the naive abstraction principle above.)

That being the case, proposals are made from time to time to retain the full abstraction principle and to try to find a way to live with (1).

What is most unlovely about (1) is that it implies, classically, the explicit contradiction

(2) $R \in R \ \& \ R \notin R$

and even people willing to live with (1) tend to consider (2) a strange bedfellow.

In fact, the passage from (1) to (2) is not hard to block; it is blocked, for example, in the many-valued logics of Łukasiewicz. But these logics are in many ways strange in other respects, and at least the finitely many-valued versions of these logics admit more subtle versions of the reasoning that leads to (2). In those logics which, as logics, are more plausible—e.g., the intuitionist logic, or the relevant logics of Anderson-Belnap—the passage from (1) to (2) cannot be blocked.

Even this, however, need not cause complete despair. Even Poincaré, no friend of formal logic, rejoiced in (2), saying that logic is not barren after all, since it begets contradictions. And reasoning close to that which produced (2) has been the very soul of 20th century logic, since closely similar reasoning has produced the interesting parts of our favourite theories and theorems—Gödel's undecidability results, Tarskian truth-theory, and the general theory of computability all share with set theory the fact that they have drawn centrally important conclusions from such reasoning.

Sometimes we contradict ourselves; this is true in our ordinary life, and it is mere prejudice which supposes that we may never do so meaningfully in our logical and mathematical life. As we have pointed out elsewhere, admitting (2) into our set theory might be of no more moment than admitting expressions that have no reasonable physical interpretation into the latest up-to-date quantum theory. The historical purpose of set theory, after all, was to provide a foundation for arithmetic and analysis, and strange theorems like (2) need not interfere with that purpose. (In the interim, of course, set theory has developed its own character and purposes, not all of which would be compatible with (2). But that strange things may go on in the upper reaches of the theoretical part of a theory, so long as our intuitions are not smashed in the parts of the theory that are close to home, is by now a commonplace. That something as strange as (2) might go on is not in general allowed, but it might as well be; at any rate, we are certainly not going to disconfirm (2) by building bigger telescopes.)

But just as (1), besides being unlovely in itself, led classically to the even more unlovely (2), so (2) leads, by classical reasoning, straight to

(3) q,

where q is any sentence whatsoever of the formal language in which we are doing set theory. Willing though we may be to contemplate (2) as

an oddity in the upper set-theoretical reaches, once we use set-theory to found garden variety mathematics there are sentences that we *definitely* want to be unprovable; e.g., there is a sentence that says that $2 + 2 = 5$, and since we want to use *that part* of our set theory for counting apples, it would be seriously distressing were that sentence provable.

We mentioned relevant logics above, and one of the nice things about them is that the inference from p & $\bar{p}$ to *arbitrary q* is blocked. So we have long considered relevant logics a possibly superior vehicle for a naive set theory with the full abstraction principle.

For almost as long, and as was first noted by Dunn and independently by Routley, we have known that the standard relevant logics R and E, even if they avoided the ultimate *reductio ad absurdum* (3) of naive set theory by the route that leads through Russell's paradox (1) and the contradiction (2), would get to (3) anyway by Curry's paradox. For a theorem scheme of these logics is the

Contraction axiom. $(p \rightarrow . p \rightarrow q) \rightarrow . p \rightarrow q$

And, as is proved by Curry and his collaborators in *Combinatory Logic,* essentially, any system which satisfies all the following conditions must be absolutely inconsistent, in the sense that every formula q of the system is a theorem of the system.

(4) All instances of the abstraction principle are theorems
(5) All instances of the contraction axiom are theorems
(6) The set of theorems is closed under replacement of equivalents
(7) The set of theorems is closed under *modus ponens* for '$\rightarrow$'.

For we may reason as follows. Let 'C' stand for '$\{x : x \in x \rightarrow . x \in x \rightarrow q\}$'. Then as an instance of the contraction axiom we have

(8) $(C \in C \rightarrow . C \in C \rightarrow q) \rightarrow . C \in C \rightarrow q$

And as an instance of the abstraction principle we have

(9) $C \in C \longleftrightarrow . C \in C \rightarrow . C \in C \rightarrow q$

By replacement of the left side of (9) for the right side in (8),

(10) $C \in C \rightarrow . C \in C \rightarrow q$

Replacing all of (10) with its equivalent by (9)

(11) $C \in C$

whence by two applications of modus ponens on (10),

(12) q

Since q may be any formula whatsoever, every formula q is provable

in this manner. (The argument above is an adaptation of the argument of volume 2 of *Combinatory Logic* for the ill effects of the Curry paradox; a slightly weaker argument, which required also that $p \rightarrow p$ be a theorem, is found in volume 1 of that work.)

That the argument to absolute inconsistency works even for E, of entailment, might dissuade even strong hearts. One originally thought that it was negation that lay behind the absolute inconsistency of naive set theory, whence one hoped that weakening classical negation priciples would at least restore absolute consistency, being willing to pay the price of simple inconsistency in order to buy the full abstraction principle. But, as we have now seen, the properties of implication—even of a relevant implication—do us in on their own.

Still, however, let us not flag. The logical trouble was with the contraction axiom, and, as many scholars have noted, it's a funny looking axiom. Curry *et al.* justify it out of their desire for something approaching the standard deduction theorem; it is this desire that makes it an axiom of the relevant logics, too. But, as Curry has also pointed out, there are versions of the deduction theorem that hold for systems *without* the contraction axiom. And so, perhaps, we should take as the logic of a naive set theory some weak relevant logic without contraction.

It is time for another blow. We have to get *very* weak. For, as we shall now see, if negation does not do us in when we drop the classical paradoxes, and if implication does not do us in when we drop the contraction axiom, we may nevertheless vary the Russell–Curry argument so that we get absolute inconsistency out of a minimal principle relating implication and *conjunction*, and which corresponds in the semantics of relevant logics to the requirement that arbitrary theories be closed under *modus ponens*. Specifically, consider the

Modus ponens axiom. $(p \rightarrow q)$ & $p \rightarrow q$

and let the set M be $\{x : x \in x \rightarrow q\}$. Then as an instance of the *modus ponens* axiom

(13) $(M \in M \rightarrow q)$ & $(M \in M) \rightarrow q$

By the abstraction principle

(14) $(M \in M \rightarrow q) \longleftrightarrow M \in M$

Replacing the left side of (14) with the right in (13)

(15) $(M \in M)$ & $(M \in M) \rightarrow q$

By the idempotence of '&' and (15)

(16) $M \in M \rightarrow q$

Replacing the whole formula (16) with the right side of (14)

   (17) $M \in M$

whence again by (16)–(17) and *modus ponens*

   (18) q

So absolute inconsistency again afflicts any system satisfying (4), (6), (7), and

   (5′) All instances of the *modus ponens* axiom are theorems, and
   (5″) Conjunction is idempotent, in the sense that $p \longleftrightarrow p \ \& \ p$

And this is an even more distressing result, since the *modus ponens* axiom, far from being funny-looking, is a direct counterpart of the fundamental logical rule.

   In short, the conclusion of our examination of Curry's paradox is discouraging in the extreme for the hopeful naive set theorist. One sought to avoid the problems posed by Russell by weakening logic, in order to save the abstraction principle. We were willing to give up the usual aversion to contradiction. We faced with equanimity the sacrifice of the deduction theorem. To continue with the project, a minimal, decent respect even for *modus ponens* must be given up as well.

   We have always held that, in these permissive days, no rule is sacrosanct—*except modus ponens*. And one may, of course, still cleave to the *rule* of *modus ponens* without the *modus ponens* axiom; indeed, Routley has conjectured that a naive set theory based on a very weak relevant logic is absolutely consistent. One might look for other escapes, too; e.g., the role of the '$\longleftrightarrow$' in the abstraction principle above might be reexamined. But the conclusion is none the less clear; unless we are prepared to give up a great deal of logic—not only of classical logic but of intuitionist and even relevant logic as well—a naive set theory is untenable.[1]

MORE ON PROPOSITIONAL IDENTITY

By Philip Hugly and Charles Sayward

LET A, B, C be sentences expressing propositions; let A be a component of C; let $C\,A/B$ be just like C except that some occurrence of A in C has been replaced by an occurrence of B; let '$=$' be a binary connective to read as 'the proposition that —— is the same proposition as the proposition that . . .'. Then

(R)   *from* $C = C\,A/B$
        infer $A = B$

is a sound rule of inference. This claim is tantamount to one made by Hugly and Sayward in [3]. In [2] James Freeman proved that if propositional identity is the same as mutual entailment, the claim is false. Here we prove the claim is true of one formal system in which propositional identity and mutual entailment are kept distinct. The system is what results from adding '$=$' to the system of tautological entailments developed by Anderson and Belnap ([1], chapter III). We shall follow them in referring to their system as $E\,f\,d\,e$. We shall refer to the enlarged system to be developed here as Id.

1. The syntax of Id. Following Anderson and Belnap we define a purely truth functional formula (p t f) as follows:

(1) A proportional variable is a p t f;
(2) if A and B are p t f s so are $-A$, $(A \,\&\, B)$, $(A \lor B)$;
(3) nothing else is a p t f.

A formula for Id is defined as follows:

(1)   if A and B are p t f s, $A \to B$ and $A = B$ are formulas;
(2)   nothing else is a formula.

Following Anderson and Belnap we shall let A, B, C . . . . range over p t f s.

2. The semantics of Id. We first make use of Anderson and Belnap's definition of an intensional lattice. An intensional lattice is a quadruple $\langle L, \leqslant, -, T \rangle$ satisfying the following conditions:

(1)   L is a distributive lattice under $\leqslant$;
(2)   for all a $\in$ L, $\bar{\bar{a}} = a$;
(3)   for all a, b $\in$ L, if a $\leqslant$ b then $\bar{b} \leqslant \bar{a}$;
(4)   T is a nonempty subset of L such that: (i) if a, b $\in$ T, then a $\land$ b $\in$ T; (ii) if a $\in$ T, a $\lor$ b $\in$ T; (iii) for all a $\in$ L, if a $\in$ T, $\bar{a} \notin$ T; (iv) for all a $\in$ L, either a $\in$ T or $\bar{a} \in$ T.

(The symbols $\land$ and $\lor$ denote the meet and join operations, respectively.

The symbol $\leqslant$ stands for any partial ordering relation. Full explanations of these concepts occur in [1], pp. 190–206).

2.1 Let $\langle L, \leqslant, -, T \rangle$ be an intensional lattice. A set D is a domain for Id iff D satisfies the following conditions:

(1)   if $a \in L$ then $\langle a, 1 \rangle \in D$;
(2)   if $\bar{a} \in L$ and $x \in D$ then $\langle \bar{a}, x, 2 \rangle \in D$;
(3)   if $a \wedge b \in L$ and if $x, y \in D$, then $\langle a \wedge b, x, y, 3 \rangle \in D$;
(4)   if $a \vee b \in L$ and if $x, y \in D$, then $\langle a \vee b, x, y, 4 \rangle \in D$;
(5)   nothing else is in D.

A domain is to be thought of as a set of propositions. The elements of a domain specified by clauses (1)–(4) are to be thought of as elementary propositions, negative propositions, conjunctive propositions and disjunctive propositions, respectively.

2.2 What has to be defined is a function that assigns propositions to A, B, $C \ldots$: elementary propositions to propositional variables; negative propositions to negations; conjunctive propositions to conjunctions; disjunctive propositions to disjunctions. Preliminary to defining such a function we need to define a function that assigns lattice elements to A, B, $C \ldots$

If $\langle L, \leqslant, -, T \rangle$ is an intensional lattice, then a function lat from L to $\{A, B, C \ldots\}$ is a lattice assignment for Id iff lat satisfies these conditions:

(1)   if A is a propositional variable, lat $(A) = a$, for some $a \in L$;
(2)   if A is $-B$, lat $(A) = \overline{\mathrm{lat}\ (B)}$;
(3)   if A is $(B \ \& \ C)$, lat (A) is lat $(B) \wedge$ lat (C);
(4)   if A is $(B \vee C)$, lat $(A) =$ lat $(B) \vee$ lat (C).

2.3 If lat is a lattice assignment for Id, a function prop is a proposition assignment for Id iff prop satisfies these conditions:

(1)   if A is a propositional variable, prop $(A) = \langle$lat $(A), 1 \rangle$;
(2)   if A is $-B$, prop $(A) = \langle$lat (A), prop $(B), 2 \rangle$;
(3)   if A is $(B \ \& \ C)$, prop $(A) = \langle$lat (A), prop (B), prop $(C), 3 \rangle$;
(4)   if A is $(B \vee C)$, prop $(A) = \langle$lat (A), prop (B), prop $(C), 4 \rangle$.

2.4 In Anderson and Belnap's semantics for $\mathrm{E}\,f\,d\,e$ the elements of L are thought of as propositions and are assigned to $A, B, C \ldots$ The partial ordering relation $\leqslant$ is thought of as entailment and is assigned to '$\rightarrow$'. It turns out that $A \rightarrow B$ is true in a model for $\mathrm{E}\,f\,d\,e$ iff lat $(A) \leqslant$ lat (B).

The relation we assign to '$\rightarrow$' cannot be a partial ordering relation. That would make propositional identity and mutual entailment the same thing. This is due to the simple fact that any partial ordering relation is antisymmetric.

What is needed here is a relation R to be assigned to '→' such that R satisfies these conditions:

> R is transitive in every domain,
> R is reflexive in every domain,
> in some domains R is not antisymmetric,
> in some domains R is not symmetric,
> in some domains R is not connected.

The following definition yields what is wanted: For any domain D, and for all x, y $\in$ D, x R y iff the first member of x bears $\leqslant$ to the first member of y.

2.5 A model for Id is a quadruple consisting of an intensional lattice, a domain, a lattice assignment and a proposition assignment.

2.6 Truth in a model and validity are defined as follows:

(1) $A \to B$ is true in a model iff prop (A) R prop (B).
(2) $A = B$ is true in a model iff prop $(A) = $ prop (B).
(3)   A formula is valid iff it is true in all models.

3. Three features of Id. According to the preceding account propositional identity and mutual entailment are distinct. In particular, this holds:

(1)   There exist models M such that $A \to B$ and $B \to A$ are true in M, while $A = B$ is false in M.

For example, these two formulas are true in every model:

$$p \to (p \vee (p \And q)), (p \vee (p \And q)) \to p.$$

While this formula is false in every model:

$$p = (p \vee (p \And q)).$$

Freeman observed that since 'p' and '$(p \vee (p \And q))$' entail each other, as do 'p' and '$(p \vee (p \And r))$', then if propositional identity is the same as mutual entailment, one could use (R) to prove '$q = r$'. But that's a pretty big 'if'. Freeman's result does not hold in Id.

Another feature of Id is this:

(2)   If $A = B$ is true in a model, so are $A \to B$ and $B \to A$.

Thus in Id propositional identity is a stronger relation than mutual entailment.

Finally observe that all the tautological entailments in Anderson and Belnap's system of tautological entailments are valid formulas in Id. That is, this is true:

(3)   If $A \to B$ is valid in E $f\,d\,e$, it is valid in Id.

The reason is that $A \to B$ is valid in E $f\,d\,e$ iff lat $(A) \leqslant$ lat (B), for all

lattice assignments lat. And, for all lattice assignments lat and proposition assignments prop, lat $(A) \leqslant$ lat (B) iff prop (A) R prop (B). Hence $A \rightarrow B$ is valid in E $f d e$ iff it is valid in Id.

4. The soundness of (R). Suppose A is a component of C; as before let $C A/B$ be the result of putting B for some occurrence of A in C. Then to show (R) is a sound rule of inference for Id is to show that if

$$C = C A/B$$

is true in a model M for Id,

$$A = B$$

is also true in M. To show this it suffices to show

(S)   if prop $(C) =$ prop $(C A/B)$ then prop $(A) =$ prop (B)

holds no matter what proposition assignment prop is.

The proof of (S) is by induction on the number of occurrences of the truth-functional connectives in C.

Basis. C is a propositional variable. Then (S) holds trivially. For C is A and $C A/B$ is B. So from the hypothesis of (S) it follows trivially that prop $(A) =$ prop (B).

Inductive step. Case 1. C is $-D$; prop $(-D) = \langle$lat $(-D)$, prop (D), 2$\rangle$; prop $(-D A/B) = \langle$lat $(-D A/B)$, prop $(D A/B)$, 2$\rangle$. From the hypothesis of (S) it follows that prop $(D) =$ prop $(D A/B)$. From the hypothesis of the induction it follows that prop $(A) =$ prop (B).

Case 2. C is $(D \& E)$. Subcase 1. $(D \& E) A/B$ is $(D A/B \& E)$. prop $(D \& E) = \langle$lat $(D \& E)$, prop (D), prop (E), 3$\rangle$; prop $(D A/B \& E) = \langle$lat $(D A/B \& E)$, prop $(D A/B)$, prop (E), 3$\rangle$. From the hypothesis of (S) it follows that prop $(D) =$ prop $(D A/B)$. From the hypothesis of the induction it follows that prop $(A) =$ prop (B).

In subcase 2, $(D \& E) A/B$ is $(D \& E A/B)$; in subcase 3 $(D \& E) A/B$ is $(D A/B \& E A/B)$. The proof that (S) holds in each of these subcases is analogous to the proof that it holds in subcase 1.

In case 3 C is $D \vee E$. The proof that (S) holds in this case is analogous to the proof given in case 2.[1]

REFERENCES

[1] Anderson, A. R. and Belnap Jr., N. D. *Entailment*, Volume I. Princeton University Press, 1975.
[2] Freeman, James B. 'A Caution on Propositional Identity', ANALYSIS 37. 4, 1977, 149–151.
[3] Hugly, Philip and Sayward, Charles. 'Prior on Propositional Identity', ANALYSIS 36. 4, 1976, 182–184.

[1] We wish to thank Professor Belnap for a helpful criticism of an ancestor to this paper.

ON THE UNIQUENESS OF THE IDENTITY RELATION

By G. H. Merrill

IN *Philosophy of Logic* (Prentice-Hall, 1970, pp. 62–63) Quine points out that the axiom

(1) $x = x$

and the axiom schema

(2) $\sim(x = y \,.\, Fx \,.\, \sim Fy)$

when added to quantification theory yield a complete proof procedure for quantification theory with identity. He then goes on to argue that

> (3) It is easy to see that these conditions permit no latitude; any two definitions of identity that meet these conditions will coincide in all their attributions of identity.

and he follows this claim with a brief proof that if two versions of identity '$x =_1 y$' and '$x =_2 y$' both satisfy (1) and (2), then it follows that

(4) $x =_1 y \equiv x =_2 y$

I have no quarrel with Quine's proof of (4), but I should like here to issue a warning against interpreting (3) in too liberal a way. His remark concerning there being *no latitude* in these conditions coupled with his claim that any two definitions (I think we should prefer 'interpretations') of identity meeting them will *coincide* may be quite misleading. For it would not be unnatural to read (3) as asserting that (1) and (2), when conjoined to standard quantification theory characterize a *unique* relation called 'identity'. But this is false, as Quine's own later remarks suffice to show.

Indeed, although (1) and (2) are the only criteria by which we may judge whether a given relation is that of identity, it does not follow (even in the wake of (4)) that any two relations satisfying these principles will be the *same* relation. To see this we need only distinguish between two interpretations of the identity sign—though for the sake of clarity we should employ *two* identity signs to avoid the appearance of paradox.

Let us interpret '$x =_1 y$' in the usual way: such a formula will be true in a model provided that that model assigns the same member of its universe to both x and y. Call the theory resulting from the addition of (1) and (2) (with '$=$' replaced by '$=_1$') to the predicate calculus the theory 'PC($=_1$)'.

Next, let us interpret '$x =_2 y$' by means of the exhaustion of lexicon described by Quine in *Philosophy of Logic*, pp. 63–4. But instead of insisting on a finite lexicon and, subsequently, a genuine *definition* of

133

'$=_2$' we may adapt Quine's idea as follows. Suppose that ϕ is an atomic formula of the predicate calculus consisting of an $n+1$-place predicate letter followed (in whatever order) by the distinct variables $\alpha_1, \ldots, \alpha_n$, and β; and assume that neither x nor y occurs in ϕ.

Define the *lexical equivalence of ϕ for x and y* to be the formula

$$\forall \alpha_1 \ldots \forall \alpha_n \, (\phi(x, \beta) \equiv \phi(y, \beta)),$$

where $\phi(x,\beta)$ (etc.) is the result of substituting x for β throughout ϕ (etc.). Then define the *lexical equivalence set of x and y* to be that set which contains as members all and only the lexical equivalence formulas for x and y. Finally, we interpret '$x =_2 y$' as true in a model just in case that model satisfies the lexical equivalence set of x and y. Note that this 'reduces' identity to the quantifiers, truth-functions, and atomic formulas but does not restrict us to a finite lexicon as Quine's suggestion does. Of course if the lexicon is not finite then there is no *definition* of '$=_2$' corresponding to this interpretation.

Call the theory resulting from the addition of (1) and (2) (with '$=$' replaced by '$=_2$') to the predicate calculus the theory 'PC($=_2$)'. Observe that PC($=_1$) and PC($=_2$) differ only notationally, and observe further that *both* the relations $=_1$ and $=_2$ satisfy (1) and (2). It is well-known that $=_1$ does. As for $=_2$, clearly the sentence

$$\forall \alpha_1 \ldots \forall \alpha_n \, (\phi(x,\beta) \equiv \phi(x,\beta))$$

will be true for *any* atomic formula ϕ containing $\alpha_1, \ldots, \alpha_n$ and β as its $n+1$ distinct variables; and so $=_2$ satisfies (1). If '$x =_2 y$' is true (in a given model), then

$$\forall \alpha_1 \ldots \forall \alpha_n \, (\phi(x,\beta) \equiv \phi(y,\beta))$$

will be satisfied for each ϕ. Hence, in particular, 'Fx' and 'Fy' will receive the same value and consequently $=_2$ will satisfy (2).

Now, however, it will be discovered that even though PC($=_1$) and PC($=_2$) have exactly the same theorems (differing only incidentally in notation), there are two very different relations of identity here. In order clearly to see this, let us consider an impoverished sublanguage of the predicate calculus—one having only the 1-place predicate A and the 2-place predicate B. (The same construction may be carried out with the full predicate calculus, as is easily seen.) In this language the lexical equivalence set for x and y is $\{$'$(Ax \equiv Ay)$', '$\forall z(Bzx \equiv Bzy)$', '$\forall z \, (Bxz \equiv Byz)$'$\}$.[1]

---

[1] Given the way I have defined 'lexical equivalence set', the set displayed here is not really the lexical equivalence set for x and y since (provided that there are denumerably many variables) that set would have denumerably many members. But most of those members would be redundant since every member of the real lexical equivalence set will be provably equivalent to a member of the displayed finite set.

Let $\mathcal{M}$ be a model of our language which ensures the following:

(a) the universe of $\mathcal{M}$ is $\{0,1\}$
(b) the extension of A in $\mathcal{M}$ is $\{0,1\}$
(c) the extension of B in $\mathcal{M}$ is $\{0,1\} \times \{0,1\}$

Consequently, $\mathcal{M}$ satisfies all formulas in the lexical equivalence set of x and y. But the sentence

(5) $\exists x\, \exists y \sim x =_1 y$

is true in $\mathcal{M}$ while the similar sentence

(6) $\exists x\, \exists y \sim x =_2 y$

is *false* in $\mathcal{M}$. (Any model differing from $\mathcal{M}$ (if at all) only in what it assigns to x and y will *still* satisfy the lexical equivalence set for x and y.) So from the point of view of $PC(=_1)$ there are *more* entities (in the universe of $\mathcal{M}$) than there are from the point of view of $PC(=_2)$. As Quine has observed, ontology is relative to language, and identity is of a piece with ontology.

Quine's derivation of (4) of course depends upon the substitution class of F in (2) containing *both* ' $=_1$ ' and ' $=_2$ ', and this condition I have violated here. Thus (3) must be interpreted as saying that when we attempt *simultaneously* to add two versions of identity to the predicate calculus, we end up with only one. But it does not follow from this that (1) and (2) define a *unique* identity relation, for there are interpretations of ' $=$ ' which satisfy (1) and (2) but do not coincide in all their attributions of identity.

Quine does point out that 'it may happen that the objects intended as values of the variables of quantification are not completely distinguishable from one another by the four predicates [of his sample language]. When this happens, [exhaustion of lexicon] fails to define genuine identity' (p. 63). And the implication here is that the failure to capture "genuine" identity is a consequence of the poverty of the language in question. Indeed, in the very next paragraph he says that 'this method of defining or simulating identity depends on finitude of the lexicon of predicates, since [the definiens] for an infinitude of predicates would never end.' However, the argument above shows that Quine's *simulation* of identity may be carried out in languages with denumerably many predicates, and it calls into question the alleged distinction between "genuine" identity and its simulacra.

In virtue of what is $=_1$ genuine identity while $=_2$ is merely a pale imitation? Is it because $=_1$ is defined via the identity relation of the *meta*language, and we know that *this* relation is genuine identity? But presumably we know this only in virtue of the fact that the metalinguistic identity relation satisfies (1) and (2). The above argument

however, shows clearly that a relation may satisfy these criteria and fail to be genuine identity. But if satisfaction of (1) and (2) is not sufficient to ensure that the relation in question is genuine identity, then surely nothing is. And it is but a short step from this to the realization that a relation of identity is "genuine" only relative to a specified meta-language containing its own identity relation. Thus while some identity relations are more genuine than others, none is the most genuine of all.

Loyola University of Chicago   © G. H. MERRILL, 1979

REPLY TO GUTTENPLAN

By JENNIFER HORNSBY

IN 'The Paratactic Account of Saying Of' (ANALYSIS 39.2, March 1979), S.D. Guttenplan wrote:

> What Hornsby needs is an account which licenses a samesaying relation between:
>   (a) 'Fingers is a crook' uttered by Ralph
> and (b) 'he is a crook' uttered by me
> but which blocks any relation between (b) and
>   (c) 'Ortcutt is a crook' uttered by me.   (p. 99)

Guttenplan is right to think that without some such account I cannot make good the extension of Davidson's treatment of indirect discourse that I proposed to cover relational sentences (see my 'Saying Of', ANALYSIS 37.2, March 1977). But I suggest that he thinks the required account is unavailable only because he ignores some features of utterances that bear on the question whether 'samesaying' relations hold between them.

Suppose that there are two particular utterances a and c described by Guttenplan's (a) and (c). Then there will be (potentially at least) utterances b' of mine that samesay with a, and utterances b'' of mine with which c samesays. But there cannot be an utterance that combines the features of b' and b''. Hence all paths from a to c that go through an utterance described by (b) are indeed blocked. For either my saying 'He is a crook' is my saying something whose content is exhausted by its representation as an open sentence, or else it is my saying something whose content is not outstripped by the content of my utterance of 'Ortcutt is a crook' (or thirdly, and irrelevantly, it does not purport to be about Ortcutt at all). In the first case my utterance samesays with a (could be used to report a), but c does not samesay with my utterance

(could not be used to report it); in the second case c samesays with my utterance, but my utterance does not samesay with a. These are things that we know on the basis of our understanding of 'say', which is the only possible touchstone for giving an account of 'samesay'.

Guttenplan's mistake was to suppose that in order to make determinate the character of an utterance it will suffice to specify a speaker and a type of sentence—as if his description (b) told us everything we needed to know about the utterances it denotes if we were to be in a position to determine what was said in them. This mistake also infects the principle (P') that he would use against me to forge links of samesaying between utterances of his type (a) and utterances of his type (c). The links connect pairs consisting of type sentences and speakers, and do not connect particular utterances having particular contents (see p. 97). I think that once we recognize that his (P') professes to state what samesays with what, but to state that without fixing enough properties of the relata of 'samesay' to fix what was said, we shall see why there is no reason to believe it. We need to know what was said, not who gave voice to which words merely, if we are to be in a position to report speech. And there can be nothing more to the question whether x samesays with y than whether x could be used to report what was said when the utterance y was made.

Corpus Christi College, Oxford   © JENNIFER HORNSBY, 1979

THE FATALISM OF 'DIODORUS CRONUS'

By ROD BERTOLET *and* WILLIAM L. ROWE

SOMEONE purporting to be named 'Diodorus Cronus' has presented an ingenious argument for fatalism whose chief interest lies in its attempt to argue directly that no one is able ever to do anything except what he in fact does ('Time, Truth, and Ability', ANALYSIS, 25.4, 1965, reprinted in Richard Taylor's *Introductory Readings in Metaphysics,* Prentice-Hall, 1978, pp. 149–53; page numbers in parentheses refer to the latter). Like most arguments for fatalism, Cronus' argument focuses on a particular example, but it's clear that if this argument works for that example, it will work for every other case in which a person does some definite act at a particular time.

Focusing on statements that are fully specific as to agent, action and time, Cronus invites us to consider an example of such a statement:

S: Stilpo walks through the Diomean Gate at t_2

which we are to assume to be true.

We may agree with Cronus that at t_2 and any later time Stilpo is not able to render S false. But how does Cronus propose to show that Stilpo was unable at t_1 (a time earlier than t_2) to render S false? He does so by arguing as follows:

> To have been able at t_1 to render S false, Stilpo would have to have been able *at t_1* to do something that would have been *logically* sufficient for the falsity of S. But nothing that he might have done *at t_1* has the least logical relevance to the truth or falsity of S. (p. 151, emphasis added)

Both statements in this argument are true, the difficulty being that the two are not logically related in the way Cronus believes. According to the first statement, Stilpo is able at t_1 to render S false just in case he's able at t_1 to do something logically sufficient for the falsity of S. Clearly this is correct. If Stilpo is unable at t_1 not to be walking through the Diomean Gate at t_2 then he is unable at t_1 to render S false. The second statement observes that Stilpo is unable to do something *at t_1* which is logically sufficient for the falsity of S. This too is true. Among the acts that Stilpo can complete *at t_1*, none is *logically* sufficient to render it true that Stilpo is not walking through the Diomean gate at t_2.

The flaw in Cronus' argument is simply this. Although he correctly observes that Stilpo is unable to do something *at t_1* that is logically sufficient for the falsity of S, he mistakenly infers from this that Stilpo is unable *at t_1* to do something logically sufficient for the falsity of S. From the fact that Stilpo can't at t_1 be *then* walking through the Diomean Gate at t_2, it simply doesn't follow that at t_1 he *then* lacks the ability to be walking through the gate at t_2. He is unable to walk through the gate later *then*, but it has not been shown that he is *then* unable to walk through the gate later. Until this latter is shown, it is not unreasonable to retain our common-sense belief that we are sometimes able to do something other than what we in fact do.

Purdue University

INDICATIVE AND COUNTERFACTUAL CONDITIONALS

By E. J. Lowe

IN his book *Counterfactuals* (Oxford, 1973), David Lewis employs the following example of a pair of sentences, one indicative and the other counterfactual:

> (1) If Oswald did not kill Kennedy, then someone else did.
> (2) If Oswald had not killed Kennedy, then someone else would have.

Lewis comments that 'the first conditional . . . is probably true, but the second may very well be false' and immediately concludes:

> Therefore there really are two different sorts of conditional; not a single conditional that can appear as indicative or as counterfactual depending on the speaker's opinion about the truth of the antecedent.[1]

While agreeing that the two sentences very probably do not have the same truth value, and are thus not logically equivalent, I shall seek to show that Lewis's conclusion does not follow. It fails to follow because one can find *another* indicative conditional which is both verbally close to (2) and logically equivalent to it. Before formulating this indicative equivalent, however, I should like to draw attention to a peculiarity of Lewis's chosen example, which makes it seem doubtful whether he could easily find an alternative example to establish this crucial conclusion.

What is peculiar about the example is that (1) has to be analysed as a *material* conditional—a relatively rare phenomenon in ordinary language. That is to say, (1) has to be interpreted as logically equivalent to

> (1′) Oswald did not kill Kennedy ⊃ someone other than Oswald killed Kennedy.

My reason for maintaining this is as follows. It seems clear that our grounds for regarding (1) as probably true are simply that it is widely believed, on very good evidence, that Kennedy was indeed killed by someone or other. In other words, we regard (1) as probably true because we take (1) to be entailed by

> (3) Someone killed Kennedy

---

and we regard (3) itself as probably true. But if (3) entails (1), then we must take these two sentences actually to be logically equivalent, because (1) evidently also entails (3), as can be seen from the fact that in no conceivable circumstances could one consistently assert both (1) and the negation of (3), that is,

(4) No one killed Kennedy.

However, if (1) and (3) are logically equivalent, then (1) must be logically equivalent to the material conditional (1′), because it is easily proved (with the resources of first order predicate logic and the laws of identity) that (1′) and (3) are logically equivalent.

This analysis is enough to establish that (1) and (2) are not logically equivalent. But it does not justify Lewis's conclusion that corresponding indicative and counterfactual conditionals require different logical treatments, unless it can be assumed that (1) really is the indicative counterpart of (2). The latter assumption will be invalid if it can be shown that some other indicative conditional exists which *is* both verbally close to (2) and logically equivalent to it, for any such conditional would obviously have prior claim to be regarded as the indicative counterpart of (2).

Now I believe that such a conditional does exist, namely the following:

(5) If Oswald has not killed Kennedy, then someone else will have.

It may be noticed that even at the lexical level the resemblance between (5) and (2) is considerably closer than that between (1) and (2), since the former pair differ merely in that 'has' is exchanged for 'had' and 'will' for 'would'. More importantly, at the syntactical level, the tense structure of (5) parallels that of (2) in a way which that of (1) does not.[2]

I offer the following argument in support of my belief that (5) is logically equivalent to (2). This argument appeals ultimately to linguistic intuition, but that is inevitable in view of the nature of the contention it is intended to uphold. Typically, (5) would be asserted by someone believing that Kennedy was destined to be murdered at some prior date and suspicious, perhaps, of Oswald's intentions, but lacking direct evidence that the assassination had occurred. Now it seems likely that such a speaker, on receiving the information that Oswald had in fact committed the crime, would instinctively amend his original assertion of (5) by asserting (2) instead. But such an amendment would not be intended by the speaker to convey any change in his previously expressed opinion: on the contrary, it would be intended as a reaffirmation of that

---

[2] Indeed, in my view, it is the tense structure of (2) which, rather than its subjunctive mood, confers upon it the modal status of a non-material conditional; the full discussion of this thesis must, however, await another occasion.

opinion, albeit within an altered framework of assumptions. This implies that (5) and (2) are logically equivalent and differ only in what they indicate as to the speaker's assumptions concerning the truth value of their common antecedent.

Lewis's chosen example, therefore, fails to establish his conclusion. But it is also doubtful whether *any* "knock-down" example could establish it. For as soon as one departs from such peculiarly simple conditionals as (1) and (2) have been revealed to be, it becomes very difficult to establish whether or not a given conditional is the counterfactual counterpart of another, indicative, conditional.[3]

University of Reading   © E. J. LOWE 1979

MOORE'S PARADOX: ONE OR TWO?

By J. N. WILLIAMS

DISCUSSIONS of what is sometimes called 'Moore's paradox' are often vitiated by a failure to notice that there are two paradoxes; not merely one in two sets of linguistic clothing. The two paradoxes are absurd, but in different ways, and accordingly require different explanations.

Moore himself seems guilty of this failure. In one place he wants to discuss why 'I went to the pictures last Tuesday but I don't believe that I did' is 'a perfectly absurd thing to say', while in another he is puzzled by the fact that ". . . it is absurd to say such a thing as 'I believe he has gone out, but he has not' . . .".

The two things which Moore alleged it would be absurd for a speaker A to say can be expressed as

 (i) 'p and it is not the case A believes that p', (e.g. 'I went to the pictures last Tuesday but I don't believe that I did').
 (ii) 'p and A believes that it is not the case that p', (e.g. 'I believe that he has gone out but he has not').

Sentences of these two forms do not express the same proposition, nor does A commit the same absurdity in uttering both.

The proposition expressed by (i) neither entails nor is entailed by that expressed by (ii). The distinction between them corresponds to the well known distinction between

(iii) 'It is not the case that A believes that p', and
(iv) 'A believes that it is not the case that p'.

The proposition expressed by (iii) neither entails nor is entailed by that expressed by (iv).

It is now uncontentious that the absurdity of A's saying either (i) or (ii) does not come from the self-contradictoriness of (i) or (ii). In both, what is expressed is possibly true, whereas what is self-contradictory is not. In both cases the absurdity lies not in what A says, but in the conjunction of what A says with his saying of what he says.

In saying 'p' one normally suggests that one believes that p, or expresses a belief that p. It is this suggestion or expression of belief that p in conjunction with the assertion, in the case of (i), that one does not believe that p or, in the case of (ii), that one believes that it is not the case that p, which gives rise to the absurdity.

But here the absurdity of (i) differs from that of (ii). For normally, it is absurd for A to assert (i) because what is conjointly expressed and asserted, i.e. a belief that p and a lack of belief that p, is *logically impossible*.

The absurdity in (ii) is of a different kind. For normally, it is absurd for A to assert (ii), not because what is conjointly expressed and asserted, i.e. a belief that p and a belief that it is not the case that p, is logically impossible, but because it is *inconsistent*.

REFERENCES

G. E. Moore, 'A Reply to My Critics' in *The Philosophy of G. E. Moore* ed. P. A. Schlipp (Evanston 1942), p. 543; 'Russell's Theory of Descriptions' in *The Philosophy of Bertrand Russell* ed. P. A. Schlipp (Evanston 1944), p. 204.

D. M. Armstrong, 'Does Knowledge Entail Belief?' *Proceedings of the Aristotelian Society* 1969–70 who, while seeing the difference in sentential form, does not indicate that the difference is anything else.

M. Black, 'Saying and Disbelieving', ANALYSIS 13 1952/53, who cites the passages from Moore without differentiating them and who says of the latter passage that it is 'Moore's latest statement of *the point*' (p. 26, my emphasis).

M. Deutscher, 'A Note on Saying and Disbelieving', ANALYSIS 25 1964/65, and 'Bonney on Saying and Disbelieving', ANALYSIS 27 1966/67, whose earlier analysis works for (ii) but not for (i) and whose later analysis works for both but fails to bring out their difference, where the later analysis is a modification of the earlier, despite the fact that the paradox accounted for in the later way is (i), whereas that accounted for in the unmodified way is (ii).

J. Hintikka, *Knowledge and Belief* (Ithica 1962), who refers to 'Moore's problem' (pp. 9, 64, 78 and 95) and who deals only with (i) (pp. 9 and 64) despite referring to passages in which Moore is concerned not only with (i) but also with (ii) (p. 64, footnote 2).

A. R. White, *Modal Thinking* (Oxford 1975) who, citing (i) and (ii) together, refers to them alike as 'the well known pragmatic paradox' and as 'this paradox' (p. 172).

University of Hull

TIME AND SELF

By EDDY M. ZEMACH

I ASSUME that a particular token of

    (1) It's two o'clock now

can be used to give valuable information. It purports to describe a fact, and it is contingently true or false.

What, then, is the statement that I make when I utter a token of (1) in answering your question, 'what time is it?'? What do I tell you about two o'clock when I tell you that it is now? The first suggestion may be that we construe (1) as having the logical form 'Fx' and say that by means of it a certain property, that of *Being Now*, is attributed to a certain temporal point, i.e., two o'clock, thus:

    (2) Now (two o'clock).

That this view is incorrect was shown, however, by many writers. Let me, therefore, mention what I believe are the main arguments against it.

(a) If *Now*, or presentness, is a property of moments of time, so are the two other temporal properties, i.e., pastness and futurity. Since two o'clock was future, is present, and will be past, it must have changed its temporal properties consecutively. But along which dimension did it change them? For a body to change e.g. its colour from red to blue there must be some dimension, i.e. time, such that on consecutive points t_1 and t_2 along it the body is red at t_1 and blue at t_2. Similarly, for the moment t_1 to change its status from future to present there must be some dimension, d, such that along it t_1 is future at one point and present at another point. But what can this dimension be? It is not time: to say that, at one o'clock, two is future, at two it is present, and at three it is past, is just to say that two o'clock is between one and three o'clock. If (1) meant, 'two o'clock is present at two o'clock' it would have been analytically true. Moreover, we would still not know whether two o'clock is now or not. So the dimension in various locations along which two o'clock is future, present and past consecutively cannot be time, and those locations on it cannot be temporal points such as one, two, and three o'clock. What dimension is it, then? Such a time-like dimension along which times move as bodies move in time is nowhere referred to in science. To postulate its existence is therefore to engage in pure fantasy and crackpot physics. In other words: if nowness is a property which is consecutively possessed by several times, there must be a location which is not itself a time which is successively occupied by the various times, such that a time, t_i, is *now* when it is at this super-temporal location. But we have no idea what such a supertemporal

dimension can be, nor do we have the slightest experimental evidence for its existence.

(b) If two o'clock can lose the property of futurity and acquire presentness, there must be some reason, a physical cause, for this change. Yet none of the properties attributable to moments of time can be considered to have brought this change about. But if physics knows of no magnitude which changes when a given moment undergoes the above change, we have every reason to suspect its reality. Entropy, of course, does *not* change with the alleged change in a given moment of time: for each moment of time, entropy is a fixed quantity.

(c) By Einstein's special theory of relativity, no two point-events are simply simultaneous or at the same place. For each point-event, any one of fully fifty percent of the universe's point-events may be deemed simultaneous with it. Instead of the ubiquitous moments of Newtonian science we have spatiotemporal point-events, each of which can be said to be either at the same time or at the same place as any given point-event. Thus even if it is given that one point-event is objectively *now*, it is meaningless to say of any other point-event that it too is objectively now. Moreover, if we say that a certain point event is objectively now, we shall have to say that it is also objectively *here* (i.e., it possesses hereness objectively, and not in relation to a given observer only). But this is surely absurd. I conclude that, because of the reasons (a)–(c) above, the first suggested reading of (1) is erroneous.

A second reading of (1) is to regard it as having the logical form '$x = y$'. Under this interpretation (1) includes two substantives referring, the one indexically and the other by name, to the same time. (1) is therefore analysed thus:

(3) Two o'clock = now.

But what is the meaning of the indexical term 'now', how can it succeed in indicating two o'clock, and why is it informative for us to learn that these two terms denote the same object? It is plausible to say that 'now' means, 'the time at which this very token of "now" is uttered'. Thus, on this reading, (1) amounts to

(4)   The time at which this very token of 'now' is uttered = two o'clock.

(4), however, is quite unhelpful as an analysis of (1). (1) was not clear because we did not know what is predicated over two o'clock when we say that it is now. But in order to learn from a token of (4) what is the time (i.e., use it as an equivalent to (1)) one must already know (e.g., by witnessing the utterance of a token of (4)) that the token of 'now' it mentions was uttered now. Otherwise (4) will give no answer at all to our question; knowing that a certain word was tokened at two

o'clock is not knowing what is the time now. I conclude that (4) cannot be used to explain the informativeness of (1) since it presupposes knowledge of, and tacit use of, the predicate 'is now'. It will not explain what is it to say of something—be it a temporal point such as two o'clock *or* an event such as the utterance of a certain word-token—that it occurs now.

In distress, we might turn to 'here', 'now' 's twin brother, for help. The informative answer 'it's Jerusalem here', which parallels (1), comes as an answer to 'what place is it?' which parallels the question 'what time is it?' The import of the spatial question is, clearly, something like this: 'which place am I located at?' The inquirer asks to be informed at which place his token of 'here' is uttered, and, consequently, he is informed where he and his answerer are located. This suggests that (1) is, similarly, informative, in that it suggests a statement like this as its import:

(5) I am located at two o'clock.

If this analysis is correct it does offer a solution to our problem. Our displeasure with (4) stemmed from the fact that there are many tokens of 'now', each referring to the time of its utterance. Thus (4) leaves open the question, which of all these tokens of 'now' occurs now, which question is then answered by a tacit use of the term 'now' itself. (5), on the other hand, does explicate what 'now' means and explains why (1) is informative. If (5) is what is meant by (1) then what (1) says is that two o'clock is *my* time, the time at which I, this unique being, am. And to know this—when do I exist—is certainly an important piece of information for me.

But then is it not the case that (5) is just false, because I exist not only at two o'clock but at other times as well? Let me start to answer this question with the obvious observation that another statement, similar to (5), is surely true. There certainly is a temporal part of me which is located at two o'clock and at no other time. But (as I have argued elsewhere[1]) the logic of our talk about material things, including persons, is such that I may consider each of my temporal parts as embodying me in my entirety. Thus one says, for example, 'I am here' and not 'here is a temporal part of me'; or 'I want to sleep' and not 'this temporal slice of me wants to sleep'. In all these sentences 'I' seems to refer to something which exists at one temporal point (e.g., two o'clock) only.

Moreover, there is an additional feature which distinguishes 'I' from most other singular terms referring to material things. Indeed, we do not say 'this temporal segment of this table is here' but rather 'this table is here', etc.; this table itself (which is both temporally and spatially extended) is talked about as if it is, all of it, embodied in each of its

[1] 'Four Ontologies', *The Journal of Philosophy*, 67 (1970): 231–247.

temporal (although not spatial) parts. Yet it is incorrect to say that there is a sense of 'this table' in which it refers to a momentary (non-temporally extended) entity. The reason is that 'this table' means, 'the entity which is the same table as (or, is table-continuous with) *this*'. The sortal concept *table* is, then, a constituent of the meaning of the indexical 'this table', and hence the referent of 'this table' is spatially and temporally extended. The situation is radically different with 'I'. As I have argued in a number of places[2], the meaning of 'I' is not restricted to any sortal. Unlike, e.g., 'this table' which applies to tables only, 'I' need not apply to objects of any particular kind. 'I', I said, means, 'the entity a part of whose career is the uttering of this token of "I" '. But what kind of entity it is and how far it extends is left open. In other words, 'this table' cannot apply to one temporal cross-section of the entity which is spatiotemporally table-continuous with *this* because the sortal concept *table* in its sense mandates that particular expansion of the reference; it governs and regulates the expansion of *this* until we reach the most remote spatiotemporal regions which are table-continuous with it. But with 'I', the situation is different. As in the case of 'this table', the indexical determines a target, a core of the object indicated, which, in the case of 'I', is the uttering of this token of 'I'. But, unlike the case of 'this table', here there is no sortal concept to govern and regulate the expansion of the reference from this core on. Therefore it is perfectly possible to use 'I' without any such expansion beyond the indicated core.

Again, if we adopt the Fregean view that senses are guises through which objects are given to the user of referring terms, we would say that the user of 'I' is aware of himself through the aspect which is the meaning of 'I'.[3] Thus the user of 'I' is aware of himself *as* the utterer of this token of 'I', and he need not be aware of himself as anything else, or in any other way. Thus there is a perfectly good sense of 'I' in which the self-referrer using it refers to an entity which is limited to the uttering of this token of 'I' only. This is the limit of what 'I' may refer to, since this is the only way in which a self-referrer *must* be aware of himself in order to use 'I'. Thus it is not wrong to say that the inquirer who asks 'what time is it?' exists at that time only. Answered by a token of (1), he is told that his interlocutor, and, therefore, he too (ignoring, for simplicity's sake, the time difference between question and answer) exists at two o'clock, and at no other time. I conclude that the analysis of (1) as (5) is correct and elucidates the importance it has in our life.

[2] 'The Unity and Indivisibility of the Self', *International Philosophical Quarterly* 10 (1970): 542–555; 'The Reference of "I" ', *Philosophical Studies* 23 (1972): 65–75.
[3] *Pace* H. N. Castañeda, who has argued, in his various articles on the logic of 'he*', that 'I' is not substitutable *salva veritate*, I suggest (using Kaplan's operator 'Dthat') that 'I' *means* 'Dthat utterer of this very (mental) token'. For a somewhat similar approach, see D. S. Clarke, 'The Addressing Function of "I" ', ANALYSIS 38.2 (1978), pp. 91–93.

How, on this view of the I as temporally unextended, can one explain the phenomena of people remembering things about their past and planning for their future? The answer is that if instead of enduring persons we only refer to instantaneous beings (which may be called 'selves') memory would be described as possession, by one self, of information about some special other selves. Those selves are those which antedate it and are *person*-continuous with it (Parfit, Lewis, Perry and others have discussed this relation; the first who treated it in any detail is, of course, Hume). Planning for the future would then be described as showing special concern for members of such a series of selves which postdate the concerned one.

To summarize: if time does not and cannot pass, if it is a mere dimension of the universe, 'now' does not denote any privileged segment of it; each spatiotemporal point in the universe is now with respect the entity whose location it is. Hence it is nonsense to say that "the past" or "the future" are not fully real; each point-event in the universe is past with respect to some selves, and future with respect to others; all are equally real. Some spatiotemporal locations host conscious beings limited to them, i.e., selves. As used by each self, 'now' refers to its time, 'here' to its place. What I have tried to show in this paper is that our ordinary time-talk makes sense only if we assume such a universe, i.e., the existence of successions of momentary selves in a spatiotemporally extended universe. In this framework, I can ask whether two o'clock is past, i.e., antedates my time, is future, i.e., postdates my time, or is present, i.e., is my time. Now, then, is the time at which I, *qua* self, exist. When I conceive of myself as a pure self, i.e., limit the information I have about myself to that which is necessary for me to call myself 'I' and nothing more, i.e., when I regard myself as having only those properties which are necessary for me in order to pick myself up as the referent of my token of 'I', then *now* is my time, my only time. I have no other.

University of Miami and
The Hebrew University of Jerusalem

EVIDENCE, STATISTICS AND RIGHTS:
A REPLY TO SIMON

By SARA ANN KETCHUM

IN a recent paper in ANALYSIS, Christine Pierce and I argued that being a member of a group, a high percentage of which are un-qualified, is not in itself a disqualification for a job.[1] For example, individual women whose physical strength is not known should not be disqualified from job J on the grounds that, for example, 80 per cent of women lack the strength necessary to perform job J. In response to our paper, Robert Simon[2] has charged that this claim is in conflict with an argument for compensatory programs such as the one offered by James Nickel.[3] Nickel argues that compensation is due to victims of past discrimination in the form of programs of special benefits. Although such programs are justified on the basis of injuries suffered, the ad-ministrative basis for the distribution of benefits will most likely be a characteristic such as race rather than proof of suffering (p. 155). White males will suffer injustice, according to Simon, 'where such preferential treatment in favour of group members requires a denial of positions to non-members' (p. 38).

I

I will agree with Simon that the argument which Nickel presents for compensatory programs would not be a particularly good argument for preferential hiring. However, the major weakness of Nickel's argu-ment is also a weakness of Simon's analysis. That is, they both assume that the connection between being black in a white-supremacist society and having suffered injustice is properly represented by the notion of statistical correlation, that the alternatives are requiring proof of suffering or distributing on the basis of a characteristic which is merely statistically correlated with being a victim of injustice. Thus, as Simon presents Nickel's argument, the relevant characteristic of the group to be compensated is that, for example, 80 per cent of its members have suffered injustice.

A comparison of two examples would show that the statistical correlation between group membership and injustice does not represent the connection assumed by defenders of compensatory programs. Imagine the following two societies:

> A. In society A there is an established pattern of white supremacy such that blacks are treated unjustly because they are black. That is, the fact

---

[1] 'Implicit Racism', ANALYSIS, 36 : 2 (1976) pp. 92, 93.
[2] 'Statistical Justifications of Discrimination', ANALYSIS, 38 : 1 (1978) pp. 37–42.
[3] 'Should Reparations Be to Individuals or to Groups?' ANALYSIS, 34 : 5 (1974) pp. 154–160.

that an individual is black is a reason, either covert or overt, for treating him or her unjustly. By some peculiar accidental circumstance, or by some weakness or failure of the system of white supremacy, 20 per cent of the blacks in this society have never suffered an injustice—they are immune not only to racism, but also to any other system of injustice prevalent in the society.

B. In society B, being black is never a social reason for being unjustly treated. There is no history or present practice of racism, black slavery, etc. However, there are practices of discrimination against three groups, and only three groups—for example, people over 65, members of a particular religious group, say, Baptists, and people who come from a particular region. Because of purely accidental demographic patterns, 80 per cent of blacks belong to one or the other of these groups and suffer the injustices levelled against these groups. But, within these groups, blacks and whites are treated exactly the same—that is a 70 year old black would be treated the same as a 70 year old white and a 30 year old black would be treated exactly the same as a 30 year old white, etc.

The argument Simon portrays would entail that we have just as much reason for assigning benefits on the basis of race in society B as in society A. And, although Nickel makes it clear that he intends to be dealing only with A, his analysis does not adequately capture the difference. But the defence of compensatory programs always takes place within a historical context, and Nickel is simply assuming that context. It would be more appropriate to assume that the characteristic which should be connected to the injustice and, hence, relevant to the distribution of compensatory benefits would be something like the following (I will call it C, for short):

C: Being an x in a y-supremacist society, where the superiority of y's is defined in terms of the inferiority of x's.

The y-supremacy might include, but not be exhaustively described by: discrimination in favour of y's against x's in hiring, access to education and other goods; a systematic ideology of the inferiority of x's and the superiority of y's; a systematic concentration of political and economic power in the hands of y's and exclusion of x's from such power, etc.

The connection between characteristic C and the property of having suffered injustice is not a merely empirical correlation. There is a logical or conceptual connection which cannot be captured by any statistical statement. The empirical question is whether or not a given society is y-supremacist and the degree and pervasiveness of that y-supremacy. Statistics are relevant as evidence that a specific society is y-supremacist, but they are not sufficient. The claim that a society is y-supremacist implies that there are practices, institutions and habits of action and thought which are (1) about the relative social position appropriate to x's and y's or about the relative capacities of x's and y's, and (2) productive of unjust inequalities between x's and y's.

Thus, I would agree with Simon that, in cases where there is a mere statistical correlation between being x and having suffered injustice it would probably not be appropriate to distribute compensatory benefits on the basis of being x. I disagree, however, with Simon's assumption that this is an appropriate description of either present-day American racism or present-day American sexism. Thus, I would not agree that the failure of this argument would constitute a point against the practice of distributing compensatory benefits on the grounds of being a black in a white-supremacist society such as our own or being female in a male-supremacist society such as our own. In such a program, one would not be distributing benefits on the basis of race or sex per se (for example, a black from a black African country might not be due any compensation for racism) but on the basis of one's past and present social status.

II

Simon seems to assume without argument that the white males who would get jobs if there were no policy of preferential treatment have a right to those jobs, and that those jobs would be, so to speak, "stolen" from them by the blacks and women who are hired on such a policy. The first thing to point out is that a policy which is preferential in intent may not be preferential in effect. Employers who are, as one should expect them to be, influenced in their thinking by the society in which they were raised and live might hire qualified blacks only if they thought they were giving preference to blacks, by the same mechanism which leads them to think of themselves as turning down unqualified blacks when they refuse to hire qualified blacks.

The problems that remain even if we assume that the policy is one that has a preferential effect and not just a preferential intent, are perhaps more serious. Let us suppose, for a particular job which requires advanced education, that under a system of blind review or an impartial merit system the proportion of white males to blacks and women hired would be roughly equivalent to their distribution in the candidate pool (that is, people with the appropriate degrees). Say the proportion is 2 per cent black and 15 per cent female. Suppose further that a policy of preferential hiring would increase those percentages to 5 per cent and 20 per cent, respectively. Thus, approximately 13 per cent of the white males who would have gotten jobs on the policy of blind hiring would not get jobs with preferential hiring. However, even with preferential hiring of this degree the proportion of white males in the profession will be several times their proportion in the population— that is, white males, who are a minority of the population, will hold a large majority of professional positions. If we are not to assume that there are vast differences in innate abilities between blacks and whites

and between men and women, then it seems reasonable to believe that the disproportionate number of white males in the applicant pool is a result of discrimination and of white male supremacy. That is, most of the difference between the percentage of white males with the appropriate training and the percentage of white males in the population is a result of past discrimination in their favour and of social and political advantages created by racism and sexism. Since, even with preferential hiring, the proportion of white males hired is considerably larger than the proportion of white males in the population, it is reasonable to assume that even such a policy of preference will grant many white males jobs they would not have received if the system had been non-discriminatory from the beginning.

Let us consider a job candidate, Mr. White, who: (1) will get job x, or some better job y, if the discrimination in favour of white males continues; (2) would have gotten job x if a policy of blind review or an unbiased application of the merit principle were used in the hiring practice and were based on present qualifications the possession of which is, in part, related to past discriminatory policies; and (3) would not have been in the applicant pool at all (that is, would not have received the necessary training) if there had been no discrimination in favour of white males in the process through which he has obtained his education. Thus, he would not have received job x, if there had been no discrimination in his favour.

Simon assumes that Mr. White has a *right* to job x (or that an injustice would be done in denying him job x), although not to job y. His assumption implies that, although Mr. White does not have a right to have discrimination in his favour continued, he does have a right to the perquisites of whatever position he has already gained from past discrimination in his favour.

Much is often made of the fact that Mr. White is not morally responsible for the injustices from which he is benefiting. But, surely, this is not sufficient to show that he has a right to benefit from these injustices. His lack of responsibility for such injustices is only relevant to the question of whether or not he should be punished for them or be required to make reparation for them. If Mr. White is prevented from benefiting from past injustices by others in his favour, it is inappropriate to claim that this constitutes his being punished or asked to make reparation. If this analysis is right, then, the suffering that Simon claims the young white males to be paying for the sins of their elders turns out to be the suffering of those who would not be able to benefit as much from past injustices as they had hoped or expected. This may be genuine and even innocent suffering, but it is not the suffering of the victim of present injustice.

Moreover, if the hiring goals are not compensatory—that is, if the

goal is to hire roughly the percentage of blacks who are in the qualified applicant pool—Mr. White's position is even shakier. In that case, the white males who do not get jobs under the system of goals will be even more likely to include those who not only would not have had the training had there never been any discrimination, but would also not have gotten the job in question if, given the present distribution of qualifications, racism were to disappear suddenly and miraculously from the hiring practice. Mr. White is then in a position of having to show that the probability and moral seriousness of his not being hired for extraneous reasons is greater than the probability and moral seriousness of some black's not being hired because of racism if the goals were not present. If we assume that there is a greater probability that white males will be unjustly treated under nonpreferential goals than that members of groups discriminated against will be unjustly treated if we allow past practices to continue, we are presupposing that members of groups discriminated against in the applicant pool are generally less qualified than white males. Unless we assume that employers will not be able to find well qualified blacks or women in proportion to their percentage in the applicant pool, nonpreferential goals should not give us any *a priori* reason for suspecting unfairness to white males.

Perhaps more important is the fact that Simon's discussion does not even touch on the major use of statistics in affirmative action law as it applies to hiring—that is, the use of statistical comparisons between the qualified applicant pool and hiring decisions as evidence of discrimination.[4] He never explains what system he is comparing preferential hiring to (although his remarks suggest that the alternative he has in mind is a perfect meritocracy run by unbiased ideal observers). Thus, he never confronts the problem of practical means of reducing discrimination. Nor does he offer any argument for the claim that the use of either preferential or nonpreferential goals will produce, for whites, more injustice than any alternative system of preventing racial discrimination would produce or allow for blacks. (He at least suggests that whatever alternative he has in mind would not involve the use of statistical goals even as evidence.)

Thus, although there may be rights which would be violated if a policy of preferential hiring were to be adopted, Simon is obscuring the issue by assuming *a priori* that the white males who would not get jobs under such a program would, in virtue of that fact, be unjustly treated. Of course there is the more radical criticism that the likely outcome of such programs would be to increase the discrimination against those groups discriminated against which are not included in the compensatory program (for example, white male second-generation immigrants). But

___

4 See Gertrude Ezorsky, 'Hiring Women Faculty', *Philosophy and Public Affairs*, 7 : 1 (Fall 1977), 82–91.

this is not an objection to the principle of distributing compensatory benefits on the basis of being a member of a group which has been discriminated against. It is rather a cautionary note about the practical application of such a principle. As such, it would not be appropriate to the conclusion Simon draws from his arguments.[5]

© SARA ANN KETCHUM 1979

State University of New York, Oswego

[5] An earlier version of this paper was presented at Hamilton College.

A DOUBT ABOUT UNIVERSAL PRESCRIPTIVISM

By JOHN IBBERSON

HOW exactly, according to Professor Hare, does a moral judgement like that expressed by the sentence (A) 'One ought to speak the truth' differ from a straightforward command like the one expressed by (B) 'Speak the truth!'? What is the difference between saying to a person that one morally ought to do a particular action, and on the other hand simply telling him to do it?

I

To answer this question Hare deploys the apparatus of 'phrastic' and 'neustic' which will be familiar to readers of his book *The Language of Morals*. For those who need reminding I will review the essentials of the distinction as briefly as possible. Sentence (B) and the indicative sentence

(C) You are going to speak the truth,

are in an obvious sense about the same thing, namely your in future speaking the truth; but they say different things about it. The indicative sentence is used to tell you that it *is* the case, while the imperative is used to tell you to *make* it the case (cf. *op. cit.* pp. 5, 17). It is just an accident of linguistic history that the words which direct attention to the thing they are both about are not the same in each. Both sentences might have contained for this purpose the common phrase 'your in future speaking the truth', adding something different at the end to show whether it was being stated to be the case ('yes') or commanded to be made the case ('please'). This would give the sentences

(D) Your in future speaking the truth, please

and (E) Your in future speaking the truth, yes.

These are synonymous with (B) and (C) respectively.

It does not really matter whether, as in these examples, the audience addressed also figures as an element in the common subject of the command and the statement. Imagine that you and I are discussing the jury of a current trial and the verdict at which they are likely to arrive. If I had access to information concerning the composition of the jury and the various bribes offered and accepted, I might tell you that the jury will find the defendant not guilty. To do this I could utter the sentence

(F) The jury's being about to find the defendant not guilty, yes.

On the other hand, I might tell you to make it the case that the jury will find the defendant not guilty, and leave it to you to decide upon the right combination of bribes and threats to employ in bringing about this state of affairs. In that case I would utter the sentence

(G) The jury's being about to find the defendant not guilty, please.

I use the 'please' in (G) to tell you to ensure, or bring about, or make the case the very same state of affairs which I use the word 'yes' in (F) to tell you is the case.

The part which is common to both imperatives and indicatives is called the 'phrastic'. In *The Language of Morals* Hare used the term 'neustic' (p. 18) as the name of that part which is different and which shows whether the speech act performed by uttering the whole sentence in a certain context is the making of a statement or the issuing of a command. Later on, in 'Meaning and Speech Acts' (*Philosophical Review* 1970, reprinted in his book *Practical Inferences*), he used the word 'tropic' for this element in order to reserve the term 'neustic' for what he calls the 'sign of subscription'. So whereas it is now the job of the tropic to show whether the sentence is in the indicative or imperative mood, it becomes the role of the neustic to indicate whether a speech act of the kind determined by the particular combination of phrastic and tropic is actually being performed at all. But since this distinction is not important for the present purpose, I will stick to the older terminology and refer to 'yes' and 'please' as the indicative and imperative neustics.

II

With the preliminary distinction between phrastics and neustics established, Hare goes on to say what he thinks are the important differences between moral judgements and ordinary commands. They centre mainly around the issue of universality. First of all, the imperative mood 'is confined, with a few exceptions which are apparent only, to the future tense' (p. 187), whereas moral judgements such as that expressed by sentence (A) apply equally to past, present, and future. If I tell someone that in certain circumstances one ought to speak the truth, and later find out that prior to this he had not done so in those

circumstances, I could justly accuse him of not having done what I said one ought to do. But if I merely *tell* him to speak the truth I cannot accuse him of previously not having done what I told him to do. For I do not tell him to have spoken the truth in the past, but to speak the truth in future. So although no action which occurred prior to the issuing of a command could count as disobedience to it, past actions can constitute violations of moral principles enunciated at a later time. Secondly, although we have singular imperatives like 'Speak the truth!' corresponding to singular 'ought'-sentences such as 'You ought to speak the truth', there are no imperatives in ordinary language corresponding to the universal 'ought'-sentence 'One ought to speak the truth.'

Having pointed out these differences he then asks how we should have to improve upon the ordinary imperative mood in order to be able to use it for the purposes now fulfilled by sentences expressing moral judgements. Since moral judgements are universal in the sense that they apply to all people at all times, the first requirement is 'to frame proper universal imperatives' so that 'we can derive from them imperative sentences in all the persons as well as in all the tenses' (p. 188). This is done by taking the phrastic of a universal indicative sentence and attaching to it the imperative neustic. The result is a sentence which can be used to 'frame a principle of complete universality, such that an action at any time whatever, done by any person whatever, could have been a breach of it' (p. 190). Thus someone's having in the past told an untruth will constitute both a counterexample to the statement I might now make by saying 'All things said being true, yes' and a violation of the command I might give by saying 'All things said being true, please'. But

> instead of the cumbrous terminology of phrastics and neustics, let us adopt the artificial [because italicized] word '*ought*'. This is to be defined as follows: if we take a proper universal indicative sentence 'All P's are Q' and split it into phrastic and neustic, 'All P's being Q, yes', and then substitute for the indicative neustic the imperative one 'All P's being Q, please', we may, instead of the latter sentence, write 'All P's *ought* to be Q'. (pp. 190–1)

With reservation he accepts the artificial word '*ought*' as a substitute for the ordinary word 'ought' in universal 'ought'-sentences such as 'One ought to speak the truth' which he says (p. 191) have the form 'All things that are said ought to be true'. His reservation (p. 192 ff.) is that although sentences with 'ought' would be able to carry out the prescriptive function of universal 'ought'-sentences, they could not, he feels, perform the descriptive function of communicating standards of value. My worry however is that they could not even be used to perform an intelligible *prescriptive* function over and above that for which we use ordinary future tensed imperative sentences.

III

For if we, so to speak, subtract the singular imperative 'Speak the truth!' from the proper universal imperative 'One *ought* to speak the truth' what do we have? The former is synonymous with (D) 'Your in future speaking the truth, please' which can be re-written as

(H) All things said in future by you being true, please.

And the universal imperative can be written as

(I)   All things said being true, please.

Clearly, the only content of (I) not exhausted by (H) is embodied in the following imperatives:

(J)   All things said at present by you being true, please
(K)   All things said in the past by you being true, please
(L)   All things said in future by everyone else being true, please
(M) All things said at present by everyone else being true, please
and (N) All things said in the past by everyone else being true, please.

When we give someone the proper universal command expressed by the sentence 'One *ought* to speak the truth' we are telling him not only to make it the case that what *he* says from now on shall be true, we are telling him also to make it the case that everything which anyone has ever said, is now saying, and ever shall say is true. What an absurd thing to tell anyone to do!

All of the commands expressed by sentences (J) through (N) are either logically or practically incapable of being obeyed. If it is not the case that everyone has done or is doing *X*, then it will be logically impossible for anyone to obey such commands as those expressed by (J), (K), (M), or (N) telling him to make it the case. On the other hand, if it already is the case then there is nothing further that anyone can do to make it the case, and hence nothing that anyone can do to obey the commands. So past and present imperatives are either logically absurd or pointless. And although it is not logically impossible to obey the future imperative expressed by (L), it is still for all practical purposes beyond the reach of any of us. How could anyone bring it about that after his death no one in any succeeding generation shall ever (knowingly) say what is false? I suppose he could destroy the planet and thereby ensure that there would be no future generations on Earth. But what about people on planets in other galaxies? How is he going to make sure that *they* only speak the truth?

In case there is any doubt at all that a universal imperative is the absurdity I make it out to be, it is necessary only to check Hare's definitions of 'phrastic' and 'neustic'. It may for instance be thought that to issue a universal imperative is to tell *everyone* to make something the

case. But this is not true, any more than to utter a universal indicative sentence is to tell everyone that something is the case. To suppose this would be to confuse the audience to which the utterance is directed with the subject about which the utterance is made. It is important in these matters to maintain a clear distinction between the people whom you are telling either that something is the case or to make something the case, and *what* you are telling them either is the case or to make the case.

Imagine that I take you aside from everyone else and say to you 'No one ever has, is now, or ever will be executed for a crime of which he is known to be innocent.' This sentence is a universal indicative paraphrasable as

> (O) No one's ever at any time being executed for a crime of which he is known to be innocent, yes,

and expresses an obvious falsehood. But the fact that it is universal does not mean that I told everyone; you were the only one I took aside and therefore the only one on that occasion to whom I asserted this falsehood. Suppose now that I retract my remark and, using the artificial word *'ought'*, say to you 'At least, no one *ought* ever to be executed for a crime of which he is known to be innocent.' I would then be uttering a proper universal imperative paraphrasable as

> (P) No one's ever at any time being executed for a crime of which he is known to be innocent, please.

Whereas previously I was telling you and you alone that something is the case, I am now telling you and you alone to make something the case—the something being the same each time. But what I am telling you to do is something which you are logically incapable of doing. There is therefore no point in telling you to do it. Yet there is obviously a point in uttering the ordinary universal 'ought'-sentence 'No one ought ever to be executed for a crime of which he is known to be innocent.'

IV

Now Hare is aware that there is something wrong with past imperatives, for he writes:

> The notion of so enriching the [imperative] mood, by producing sentences (such as past imperatives) that could have no use in our language, may well excite suspicion. It is obvious why we never command things to happen in the past; and therefore it might be said that a past imperative would be meaningless. I am not concerned to deny this—for in a sense an expression *is* meaningless if it could have no possible use; but

> nevertheless it will be seen that these sentences do have a function in my analysis, and therefore I must ask the reader to put up with them. There is perhaps an analogy with the use of imaginary numbers in mathematics. (p. 188)

I am not sure exactly what the analogy with imaginary numbers is supposed to be, but it is clearly a consequence of his analysis that so far from never commanding things to have happened in the past, we do precisely this whenever we affirm a universal moral principle. For according to the analysis, part (if not all) of doing so is issuing a proper universal imperative. Yet it seems perfectly bizarre to suppose that telling someone to make something to have been the case is any part at all of telling him that it ought to be the case. On p. 195 a complaint is mentioned, and rejected, to the effect that

> *'ought'*-sentences would somehow lack the authority which attaches to 'ought'-sentences of ordinary language. When I used *'ought'*, I should be only *telling* people to do a certain kind of act; but when in ordinary language I say that they ought to do a certain kind of act, it is not just that I am telling them; . . .

Again, what is objectionable is not so much that you should be merely telling people to do something. This is certainly bad enough. But it is *what* you would be telling them to do that is so incredible. To speak the truth in future is something which it is reasonable to tell people to do. It is hardly reasonable to tell people to make it the case that they and everyone else have always in the past spoken the truth—especially when we know perfectly well that this is in fact not the case. Yet if Hare is right this is exactly what we are doing in asserting a moral judgement. It would seem then that if we wish to avoid speaking nonsense we should abjure moral judgements altogether and restrict ourselves to plain imperative sentences of the future tense.

This is an undesirable conclusion, but one which is forced upon us if we accept the premise that moral judgements are universal prescriptions. They are indeed temporally universal in the sense that actions occurring before, after, or simultaneously with their assertion can count as violations of them. Hare requires them also to be prescriptive so that they can entail singular future tensed imperatives; in other words, so that accepting the universal 'ought'-statement expressed by 'One ought to do X' logically commits one to accepting the singular command expressed by 'Let me in future do X' (or 'My in future doing X, please'). The trouble is that universal prescriptions entail much more than this, and the more which they entail is something no one would ever wish to say, let alone commit himself to undertaking.

Oriel College, Oxford   © JOHN IBBERSON 1979

RUSSELL AND FREGE AGAIN

By P. T. Geach

IN a previous article (ANALYSIS 38.2, June 1978) Blackburn and Code maintained that in all important respects Russell's theory of denoting in *The Principles of Mathematics* coincides with Frege's theory of *Sinn* and *Bedeutung*. In their recent article (ANALYSIS 38.4, October 1978) they disclaim any intention to have described Russell's earlier use of 'denoting phrase'; but by declining to discuss how Russell at that date actually talked about denoting, they disqualify themselves from expressing any view on what his earlier theory of denoting was. They are misguided in citing 'Apollo' as a counterexample to my own thesis that for Russell proper names are not denoting phrases: on the relevant page of 'On Denoting', Russell spells it out that he regards 'Apollo' as short for 'the sun-god', so that he is treating it precisely *not* as a proper name, but as an abbreviation for a denoting phrase. With this I must leave the historical issues between these authors and myself to the judgment of readers of ANALYSIS.

I have argued that in cases where Frege would say we recognize over again the same *Sinn* of a proper name, we can indeed recognize a repeated use of the name with a sustained and sharable intention of a common reference. (An intended reference e.g. to a man is not a kind of reference to a man, any more than a projected murder is a kind of murder; but like projects of murder, shared intentions to make a reference are often successful.) With this sort of elucidation, then, I indeed proposed to explain 'recognition of a certain *Sinn* expressed by a proper name' as replaceable by 'recognition of a certain use of a proper name'. This however does not commit me to a simple equation between '*Sinn* expressed by . . .' and 'use of . . .', or between Frege's '*einen Sinn ausdrucken*' and 'express a use'; to hold me thus committed is a clear example of the cancelling-out fallacy that I censured in *Reference and Generality*. However, I ought to have seen that this fallacy was one that unwary readers were likely to fall into, and to have given an explicit warning against it.

Neither is it relevant for Blackburn and Code to protest that the sense, i.e. *Sinn*, of a sentence cannot be equated with the use of a sentence. I discussed only our discriminative recognition of the *Sinn* of proper names. When the same *Sinn* is to be recognized in two different sentences is clearly a very difficult question; I did not raise it, and expressed no view about it, even by implication. I do not think my failure to do so casts doubt on my argument that we can and do recognize when our fellows use an equiform proper name with one or another *intended* reference, independently of any cognitive contact with the actual objects

(if any) that are being referred to; and that this recognition is precisely a recognition of what Frege called the *Sinn* of proper names.

University of Leeds   © P. T. GEACH 1979

GEACH AGAIN

By SIMON BLACKBURN *and* ALAN CODE

1. WE have said all along that *PoM* and Frege *differ* in certain important respects. In ANALYSIS 38.2, March 1978, we tried to show that differences which might seem relevant to pp. 48–50 of OD do not, in fact, matter to the argument on those pages.

2. While not attempting to describe *PoM*'s usage of 'denoting phrase' we did manage to say quite a bit about how that book treats the denoting phrases under scrutiny on pp. 48–50 of OD. We, at least, have not 'declined to discuss' this.

3. Our position requires no more than the fact that 'Apollo' is syntactically a proper name regardless of its semantics.

4. Last time Geach said on p. 205: 'Now these recognizable, identifiable and discriminable, *uses* of names simply are what Frege meant by *senses* of names. . . .' We are blithe about interpreting this as asserting that Frege held that the sense of a name is a use. Had Professor Geach said only what he now wishes he had said there would have been a 'cancelling-out' fallacy, but as he didn't, there wasn't.

5. The problem which, according to us, Russell found in Frege's theory, is not met by describing ways in which a name can have a use or sense regardless of what it names or of whether it names anything. Obviously Russell himself thought this of most ordinary (syntactically identified) names. It is the switch of reference to these senses which, according to us, Russell attacks. He thinks that if you switch reference from context to context you incur a logical obligation, and he thinks that the obligation cannot be met, given the kinds of way in which senses must be described. We think that the first claim ought to be obvious to any serious semantic theorist. And we think the second claim is plausible enough to earn for Russell's pages a respect which they are not normally paid. This was the content of our paper.

Pembroke College, Oxford   © SIMON BLACKBURN *and*
University of California, Berkeley   ALAN CODE

NOTES

The ANALYSIS Committee consists of: Chairman, P. T. Geach; Secretary, J. H. Benson; T. R. Baldwin, A. E. Fisher, Andrew Harrison, R. F. Holland, Hidé Ishiguro, J. Kemp, J. H. McDowell, Bernard Mayo, D. H. Mellor, R. G. Swinburne, A. R. White, C. J. F. Williams, Peter Winch. This committee is responsible for appointing and advising the Editor and for the general policy of the paper.

SUBSCRIPTIONS. The subscription to ANALYSIS for institutions is £5.00 (inland), £6.00 (overseas), $15.00 (U.S.A. and Canada); for individuals £4.00 (inland), £4.80 (overseas), $12.00 (U.S.A. and Canada). Each volume comprises four numbers, three of 48 pages and one of 64 pages, appearing within the academic year—in October, January, March and June. Orders should be sent to Basil Blackwell, 108 Cowley Road, Oxford OX4 1JF; or placed with any bookseller.

CONTRIBUTIONS. **Articles submitted for publication should be addressed to Mr. Christopher Kirwan, Exeter College, Oxford OX1 3DP.** Contributors are asked to note the following requirements.

Articles should **normally** not exceed 3,000 words in length. **Occasionally,** however, longer contributions can be accepted;

They must be type written in double spacing on one side of the paper only;

Footnotes should be kept to a minimum and wherever possible avoided altogether;

Single quotation marks should normally be used, except for purposes of internal quotation and "scare" quotes.

Discussion papers should be sent in as soon as possible after the appearance of the article to which they refer.

It is regretted that owing to increased postal charges it is no longer possible to return typescripts unless the following instructions are followed:

Contributors in the United Kingdom should enclose a stamped addressed envelope of suitable size; if immediate acknowledgement is required, a stamped postcard should also be enclosed.

Overseas contributors who wish to have their MSS. returned should send an envelope **and international reply coupons** of the requisite value, whether for air or surface mail.

Galley proofs of accepted articles will be sent to authors for correction, together with information about offprints. Typescripts will be retained by the Editor on the assumption that authors have kept their own copies.

The copyright of articles printed in ANALYSIS remains the property of the author, but contributors are strongly advised, in their own interest, to consult the Editor before consenting to the reprinting of their articles.

ANALYSIS

Edited by

CHRISTOPHER KIRWAN

CONTENTS

ANALYSIS

Edited by Christopher Kirwan

Volume 39

1979

INDEX

UNIVERSAL AND PAST-TENSE PRESCRIPTIONS:
A REPLY TO MR. IBBERSON

By R. M. HARE

I am extremely grateful to Mr. John Ibberson ('A doubt about universal prescriptivism', ANALYSIS 39.3, June 1979) for rescuing the last chapter of my *The Language of Morals* from the oblivion in which it has been lying since publication (the rest of the book by contrast has achieved too much notoriety). He has, moreover, taken pains to report what I say accurately.[1] I do not wish to maintain all the theses suggested in that chapter.[2] It was something of a try-out, on which I hoped that somebody more skilled than myself in formal logic might improve; but so far as I know nobody has tried. It was also tentative in another way: it was avowedly an artificial model which made no claim to represent accurately all the features of our ordinary value-words. I said of these:

> They are indeed so varied in their uses and so subtly flexible that any artificial construction is bound to be a travesty of them (*LM* p. 180).

So I am not going to complain when Ibberson shows precisely this. All I was trying to do was to display rather crudely the main pattern of my proposed account of them; and, in order to make it plainer, I took, and openly paraded, some liberties which I am still ready to defend, at least as expository devices.

The liberty on which Ibberson concentrates his attack is that of 'enriching' the imperative mood in order to make it possible to frame in it properly universal prescriptions. These are what I need if I am to use the mood in constructing an analytical model of value-judgments, which are, on my view, a kind of universal prescriptions. In order to achieve this, I need to have a kind of imperative which, like a universal indicative, when combined with the appropriate minor premisses

---

[1] Unlike Mrs. Vera Peetz, who at the beginning of her paper 'Imperative Inference' (ANALYSIS 39.2, March 1979) attributes to me the view that we can draw an imperative conclusion from premisses which are all indicatives, and says that I share this view with Professor Geach. She cites '*Language of Morals*' without chapter or verse. If she had verified the reference she would have discovered the opposite view advanced on p. 28, though qualifications to this tentatively expressed thesis are needed (see my Preface and *Philosophy* 52 (1977), where the qualifications in respect of 'ought'-statements could be extended to imperatives).

[2] For example, the definition of 'better than' in terms of 'ought' given on p. 184 produces the paradox that *A* could be a better *X* than *B* and also *B* be a better *X* than *A* (for it is possible consistently to say 'If one is choosing an *X*, then, if one chooses *A*, one ought to choose *B*, and if one chooses *B*, one ought to choose *A*—i.e. one ought to choose either both or neither').
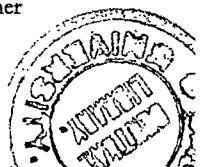
entails singular sentences[1] in any person and any tense. Ibberson objects, reasonably, that we do not have past-tense commands in ordinary language, and maintains further that we could not give them a use if we tried to introduce them. That I felt the same doubts myself is evident from *LM* p. 188 and elsewhere, and from my earlier article from which many of these ideas come. ('Imperative Sentences', *Mind* 58 (1949), p. 25, reprinted in my *Practical Inferences*, pp. 5f). But he would perhaps have been more indulgent if he had understood my motives for taking this liberty, which I will now try to explain more fully.

It is essential to the meaning of moral and other evaluative principles that we can meaningfully point to past breaches of them. We do this whenever, in teaching such a principle, we use a past breach in order to impart a lesson. Thus, in order to teach a learner how to use the clutch correctly when the car has an old-fashioned gear-shift, I may say 'You ought to have let the clutch in more gently' (cf. *LM* pp. 152, 155–7). I was impressed by the fact that, in such cases, instead of saying that we were teaching him how he *ought to* drive, we could almost equivalently say that we were teaching him how *to* drive, and in particular how *to* treat the clutch; the principle that we were teaching could, indeed, be expressed in the imperative: 'Always let the clutch in gently'.[2] If I am to use a past breach of this principle in order to give instruction in the principle, it certainly looks as if the learner, when he let in the clutch too roughly, was breaking the *same* principle which I am still, now, trying to teach him, viz. the principle which could, now, be expressed by 'Always let the clutch in gently' or 'One ought always to let the clutch in gently'. The latter is perhaps the preferable form, because the former might, as Ibberson's argument well shows, lead to trouble. It might be taken to be an ordinary less-than-universal imperative, and to have no application to past instances. But all the same it is a very natural expression for what we are trying to teach; and in teaching it we are clearly allowed by ordinary usage to point to past breaches in order to make clear to the learner exactly what the principle requires him to do or avoid.

We might compare the notice in the railway carriage which says 'No Smoking' (cf. *LM* pp. 175 ff). Although I was well aware that this does not express a properly universal command, it might be held to

---

[1] In *LM* I used the slovenly pre-Strawsonian terminology (still surprisingly current) which did not distinguish between sentences and statements or commands. Strictly, sentences do not entail one another. I also used the term 'command' in its generic sense, as the grammarians do, where I would now prefer the term of art 'imperation'; see *LM*, p. 4 and my *Practical Inferences* (London, 1971), p. 56. I may occasionally in this paper use the old terms for ease of comparison.

[2] I follow Fowler (*Modern English Usage* s.v. 'Stops') rather than Mr. Ibberson and most philosophical writers on this subject, and use no exclamation mark after imperatives. That these writers really agree that this is correct is shown by the fact that when they have occasion to *use* an imperative in their text, as opposed to mentioning it, they never commit this solecism (cf. Ibberson's p. 154 line 4).

entail *some* past-tense imperatives, on the ground that a person who is apprehended for breaking the 'No Smoking' rule cannot escape retribution by pleading that nobody can now with point express the rule that he broke. Such a plea, if set out more fully, might run: 'I admit that I smoked in the carriage five minutes ago; but you cannot tell me not to smoke five minutes ago; so the rule 'No Smoking' which you are *now* expressing and trying to enforce cannot apply to the time when I smoked, because if it did that it would have to entail past imperatives, which can have no use. So I broke no rule which you can now with point utter; so you must let me off'.

Of course there are various more or less cumbrous ways of seeking to avoid this difficulty. We might try saying that the notice in the railway carriage constantly changes its meaning from second to second. At 5 p.m. it means 'Do not smoke in this carriage after 5 p.m.'; but at 5.5 p.m. it has changed its meaning to 'Do not smoke . . . after 5.5 p.m.'. The objection to this is that we do want to say that breaches at different times are breaches of the same rule (instruction in rules would, as we have seen, be difficult if this were not so). I do not think that the difficulty can be avoided by changing the time-reference '5 p.m.' to the token-reflexive expression 'now'; for then, in saying 'Do not smoke . . . after now', we would be uttering a rule containing a constantly-changing reference, which would be just as bad.

A more promising move is to say that the notice always means the same, namely what the person meant who first put it up. *He* was referring to all times in the future and not to any past times. What the offender is punished for is breaking the future-tense commands given in the past by this august person.[1] However, even now a difficulty remains. When *we, now,* require that the commands of the august person be obeyed, are we uttering constantly-changing requirements? That is to say, are we at 5 p.m. requiring everybody to obey his rules from 5 p.m. onwards, but at 5.5 p.m. requiring people to obey them only from 5.5 p.m., and so on? If the high-ups in British Rail are asked *now* to say what their requirements are, are they forced to choose between saying something quite pointless, viz. 'Obey in the past the rules made by the august person in the past' as well as something pointful, viz. 'Obey in the future the rules made by the august person in the past', and saying only the latter? But if they say only the latter, they cannot call past breaches breaches of their present requirements.

We might try saying that their present requirement is 'Punish all those who have in the past broken rules made in the past by the august person'. But this creates the same difficulty: is *this* a constantly-changing

---

[1] In the case of moral judgement God might serve as the august person—in which case we perhaps ought, as Professor Anscombe seems to wish, to stop making them if we do not believe in him: cf. *Philosophy* 33 (1958), p. 18. His eternity might solve the problem.

requirement, and can officials who have failed to prosecute offenders plead that they have not broken any present requirement of the high-ups?

It is natural, though for reasons which Ibberson has given logically dubious, to cut this Gordian knot by allowing there to be a rule 'No Smoking' which does *not* change its meaning or reference from moment to moment. I agree with Mr. Mackie (*Ethics: Inventing Right and Wrong,* Penguin 1977, p. 42) that there cannot be 'objective prescriptivity', and would indeed go further than him and claim that the notion of objective prescriptions is incoherent. This, however, is not because prescriptions cannot be universal but because they cannot be factual. Indeed, it is possible that even our ordinary usage countenances properly *universal* prescriptions: prescriptions whose application extends to all times and all persons characterized by the features specified in the prescription. It is my view that 'ought'-sentences express such prescriptions; and I am inclined to think that the ordinary man would not object overmuch if it were claimed that the 'No Smoking' notice expresses a prescription which, though not *fully* universal (for it applies only to this carriage, cf. *LM* p. 177) is at least universal enough to entail some past imperatives.

What are we to make of these? My analogy with imaginary numbers in mathematics (*LM* p. 188) was intended to suggest that the uses permitted by ordinary language are extensible for technical purposes in the way that I am proposing. We learnt the use of the number-words in counting; and we cannot use the square root of minus one and its multiples in counting; yet mathematicians and physicists have found a use for it in their theories. We should not be put off, by the fact that past-tense imperatives do not by themselves have a use in commanding, from adopting an analysis of value-judgments which produces them as a by-product. Nor is it necessary here to go into the vexed question of whether the impossibility of altering the past, which is the reason why past-tense imperatives lack a use, is a logical or conceptual impossibility or a causal or physical impossibility. It is enough to say that we *cannot* obey past-tense commands, and so we can simply ignore, in deciding what to do, the past-tense commands which are entailed by universal prescriptions. The only bearings they have on our conduct are those already mentioned: by noting that certain past actions were breaches of them, we can be helped to learn the universal prescriptions which entail them, and these in turn may guide our future conduct (cf. *LM* p. 129); and we can punish people for past breaches of the same rules as are still in force. Even *new* retroactive criminal legislation is merely repre-hensible, and not logically impossible.

Mr. Ibberson's other difficulty, with imperatives in persons other than the second, is more easily dealt with. He rightly distinguishes between the person or thing referred to by the grammatical subject of

an imperative, and the person to whom the command is addressed. There is also a third kind of person who might in another sense be called 'the person commanded', namely the person whom the commander holds responsible for executing the command. If the commanding officer says to the adjutant 'The leading vehicle will (i.e. is to) move off at 0700 hrs.', then all three 'subjects of the command' are distinct, viz. the adjutant, to whom it is addressed, the leading vehicle, to which the grammatical subject refers, and the driver of the leading vehicle and his immediate superiors, who are responsible for executing the command. These distinctions were not brought out by the simple cases considered in the introductory chapter of *LM;* but the difficulty which troubles Mr. Ibberson is readily diagnosed and removed if we distinguish, much as Mr. Kenny and his predecessors have done, between *directives,* which place the responsibility on the addressee, and *fiats,* which do not (ANALYSIS 26.3, January, 1966, p. 68). If we then say that universal prescriptions are fiats, we shall be freed from the absurdity of supposing that when I utter one, I am commanding the person addressed to make it the case that all the singular commands which my universal prescription entails are obeyed. The most I am doing is uttering a fiat which anybody who, in the circumstances specified in the prescription, did not do the action specified would be failing to satisfy.

To know this, and to know what are the specified circumstances and the specified action, is to understand the fiat. There will of course be many (the vast majority) of those in such circumstances who never come to hear of my fiat; but they are nevertheless in breach of it, because they do not satisfy it. And moral judgments seem to me to be indeed like this: if I say 'One ought never to tell lies', anybody who ever tells or has told a lie has erred against the principle which I thus express, or, as I would equivalently put it, failed to satisfy my universal prescription.

Corpus Christi College, Oxford   © R. M. HARE 1979

PARADOX AND SEMANTICAL CORRECTNESS

By PHILIP HUGLY *and* CHARLES SAYWARD

IN [1], [2] and [3] R.L. Martin has developed a theory which is supposed to solve the semantic paradoxes. The theory not only fails as a solution to the paradoxes, it is inconsistent as well.

I

In [1] the following ideas are presented: Martin claims that 'This very sentence is false' is not 'semantically correct', and that all semantically incorrect sentences lack a truth value. Martin does not define semantical correctness but he gives a procedure for determining semantical correctness. This procedure essentially involves determining whether what is indicated by the subject of the sentence is included within the range of the predicate, which is the set of things to which the predicate is significantly applicable.

We are asked to consider what happens when we apply the procedure to 'This very sentence is false', which he refers to by '(a)' ([1], p.299):

> Taking (a) . . . as a self referential sentence type, we next ask, according to the procedure, whether the demonstrative reference of the subject is included in the range of applicability of the predicate, "is false." But the demonstrative reference of (a) is (a) itself; hence, before we can proceed, we must determine whether (a) is semantically correct. Here of course our inquiry breaks down; (a) can pass the test devised for self referential sentences only after it passes that test. Granted, the procedure does not yield the result that (a) is semantically incorrect; what we see, rather, is that (a) can never reasonably be ruled semantically correct.

It is essential to this argument that the range of 'false' is the set of semantically correct sentences. If the range of 'false' is the set of grammatical sentences Martin's test *does* establish the semantical correctness of (a).

Here is what Martin says in defence of his claim about the range of 'false' ([1], p. 299):

> I have already urged that we accept the characterization of the class of semantically correct sentences as those sentences which are either true or false. It seems equivalent to this characterization to state that the range of applicability of the predicate "is false" and its category mate "is true" is the class of semantically correct sentences.

These two propositions are being held equivalent:

(1) The class of semantically correct sentences = the class of sentences true or false.

(2) The range of 'is false' = the range of 'is true' = the class of semantically correct sentences.

The equivalence claim needs argument. Why could one not consistently hold that 'true' or 'false' are significantly predicable of sentences of English outside the class of truth-valued sentences of English, while at the same time adhering to (1)?

At any rate the following thesis is central to Martin's theory:

(M1) R ('is true') = R ('is false') = {x: x is a sentence of English that is true or false}.

Now we shall argue that this thesis is inconsistent with other theses which are central to the theory.

These other theses emerge in Martin's latest paper ([3], p. 466; we correct two minor misprints):

> I assume ... that the sortal range of a complex n-place predicate (that is, a molecular or quantified formula containing free occurrences of n distinct variables) is to be construed as some function of the sortal ranges of the constituent predicates. The question is, what function? As a first step we shall need to say what is the sortal range of $\neg Fx$ given the sortal range, R(F), of F, and what is the sortal range of '$Fx \vee Gx$' given the sortal range of F and G; and for present purposes that will suffice. The sense of negation appropriate to a category scheme is, I take it, choice negation—so we construe R($\neg Fx$) as the same as R(F). As for disjunction, I think an argument against taking *union* as our function should suffice as a case for choosing intersection; and treating R($Fx \vee Gx$) as R(F)$\cap$R(G) amounts to construing the connective *weakly*: the *disjunct makes the disjunction *.[1] The argument against union as the projecting function is simply that, in connection with the usual definition of *conjunction* in terms of negation and disjunction, it yields the unwanted result that a conjunction A & B can be true even when one conjunct is *. For example:

$$-(-(\text{2 is even}) \vee -(\text{2 is green}))$$
$$\text{T F} \qquad \text{T} \quad \text{F*} \qquad *$$

Thus these two theses are also part of Martin's theory.

(M2) For any predicate P, R ($\ulcorner \neg P \urcorner$) = R(P).
(M3) For any predicates P and Q, R($\ulcorner P \vee Q \urcorner$) = R(P)$\cap$R(Q).

In addition, Martin accepts these two theses:

(M4) The extension of a predicate is a subset of its range.
(M5) 'Saturday flies' is not true or false.

The theses (M1)—(M5) form an inconsistent set. Here's the proof:

(1) 'Saturday flies' $\in$ R ('is not true or false'). By (M4) and (M5).
(2) 'Saturday flies' $\in$ R ('is true or false'). (1), by (M2).
(3) 'Saturday flies' $\in$ R ('is true'). (2), by (M3).

---

[1] Martin uses '*' to indicate that a sentence is without truth-value.

(4) 'Saturday flies' $\in$ {x : x is a sentence of English that is true or false}. (3), by (M1).

(5) 'Saturday flies' is true or false. (4), by set theory.

If Martin's theory is to be salvaged at least one of the theses (M1)–(M5) have to be given up. (M1), (M4) and (M5) are crucial to him. That leaves (M2) and (M3). Could either of these be given up?

If (M2) is given up then the range of 'is not true' would not be the set of truth-valued sentences of English. Thus it would not be the set of sentences of English which are semantically correct. In that case Martin's account would not apply to 'this very sentence is not true'.

The only viable alternative to (M3) that we can think of is

(6) $R(\ulcorner P \vee Q \urcorner) = R(P) \cup R(Q)$.

We have already seen Martin cite one problem with (6). Here's another. Given (M1) and (M2), (6) yields

(7) R ('is not true or false') = R ('is true').

But Martin wants to say 'Saturday flies' is in the first set and not in the second.

II

Martin also claims to have solved the Grelling paradox. The idea central to his proposed solution is this ([3], p. 469):

> Given that each 1-place predicate P has a sortal range of $R(P)$ of objects of which it is true or false, I propose that the pairs in the sortal range of 'is true of' be those whose 1st element is a 1-place predicate and whose 2nd element belongs to the sortal range of the first element.

Let 'E' abbreviate 'the expression "is an English expression that is not true of itself"'. Then consider

(b) E is not true of E.

This has a truth value only if the pair $\langle E, E \rangle$ belongs to the sortal range of 'is true of'. So does E belong to $R(E)$?

> We know . . . that the sortal range of E is the same as it would be without a negation; assuming a parallel with the sortal range of 'is true of', the sortal range of E consists of exactly those English predicates that have a truth value when applied to themselves. But that is just the question, once again, whether E belongs to $R(E)$. ([3], p. 470).

The circle is supposed to keep us from any basis for assigning a truth-value to (b).

Crucial to this argument is

(M6) $\langle x, y \rangle \in R$ ('is true of') if and only if x is a 1-place predicate of English and $y \in R(x)$.

But it turns out that (M6) is inconsistent with (M2)–(M5). For consider:

(8) 'flies' is not true or false of Saturday. By (M5).

(9) ⟨'flies', Saturday⟩ ∈ R ('is not true or false of'). (8), by (M4).

(10) ⟨'flies', Saturday⟩ ∈ R ('is true of'). (9), by (M2) and (M3).

(11) Saturday ∈ R ('flies'). (10), by (M6).

(12) 'flies' is true or false of Saturday. (11) (since $x \in R(P)$ if P is true or false of x).

III

Martin's proposed solution to the Liar depends on (M1). His proposed solution of the Grelling depends on (M6). Nowhere do we find an argument for either principle. This fact alone renders it doubtful whether the proposed solutions are genuine solutions. But worse than this is the fact that (M1) is also inconsistent with (M2)–(M5), and the fact that (M6) is also inconsistent with (M2)–(M5). Then there is also the fact that none of the principles (M2)–(M5) can easily be given up. So very little can be said for Martin's theory.

© PHILIP HUGLY *and* CHARLES SAYWARD 1979
University of Nebraska—Lincoln

REFERENCES

[1] Martin, R. L. 'Toward a Solution to the Liar Paradox', *Philosophical Review*, LXXVI (1967), 279–311.
[2] Martin, R. L. 'On Grelling's Paradox', *Philosophical Review*, LXXVI (1968), 321–330.
[3] Martin, R. L. 'On a Puzzling Classical Validity', *Philosophical Review*, LXXXVI (1977), 454–473.

REPLY TO HUGLY AND SAYWARD

By ROBERT L. MARTIN

HUGLY and Sayward (ANALYSIS 39.4, October 1979) present three criticisms of my work on the Liar paradox. (I will confine my comments to the Liar; the discussion is easily applicable to the case of the Grelling paradox). They claim:

I it is doubtful whether I have offered a genuine solution, because, they claim further, I have no argument for my

---

¹ By a *bivalent sentence* I understand one that has a truth-value, of which I assume throughout that there are exactly two, true and false; however, as will emerge, I propose to treat 'bivalent' as a semantical primitive.

principle M1. (M1 states that the sortal range of 'true' and 'false' is the set of bivalent sentences[1]).

II M1 is inconsistent with four other principles, M2—M5, which I also hold.

III none of M2 — M5 can easily be given up.

I will comment on these in turn.

I

I believe the category solution, as I call it, is at least a genuine contender. I presented an argument for M1 (in CSL, p. 92) which I will rehearse in a moment, but first I want to say something about solutions to the semantic paradoxes. I consider my own early views on the matter simplistic, as when I wrote: 'A solution to the Liar . . . consists in showing that one's way of removing the contradiction is natural and intuitive.' (TSLP, pp. 292–3) I now suspect that all solutions to the paradoxes will be *un*natural and *counter*intuitive in some way or other— my evidence is simply that the contradiction is derivable from *very* natural and deeply held semantic principles. I am now inclined to the view that intuitive semantics is in fact inconsistent, and that 'solutions' are really proposals for a reformed concept of truth. Of course, arguments for particular principles are still needed, but they are unlikely to be conclusive; one should keep in mind the advantages and disadvantages of the solution as a whole.

Hugly and Sayward are right to question the argument for M1 that I gave in TSLP. I had argued that sortally incorrect sentences (terminology I now prefer to 'semantically incorrect') are without truth-value, and suggested, confusedly, that this was somehow equivalent to the view (M1) that the sortal range of 'true' and 'false' is restricted to sentences with truth-value. One has only to raise the question to see that there is no such equivalence: one can hold that ' "Saturday flies" is true' is bivalent (and false) while denying that 'Saturday flies' is bivalent.

The idea I was groping for is, I believe, as follows: *Suppose* one has settled on a semantic, presuppositional analysis of the sortal-incorrectness phenomenon. That is, suppose one holds that sortally incorrect sentences lack truth-value (as opposed to: are ungrammatical, or are grammatical and bivalent but contextually deviant[1]), and that the predicates of natural language exhibit a presuppositional tree structure whereby, if a 'category predicate' is *false* of an object, the 'family-member' predicates falling under the category predicate are

[1] Cf. R. H. Thomason's discussion of these alternatives, and his arguments in favour of the semantical approach, in 'A Semantic Theory of Sortal Incorrectness,' *Journal of Philosophical Logic* 1 (1972) pp. 209–258, esp. pp. 229 ff.

non-bivalent of the object. The picture, adapted to a particular example, is as follows: .

has a colour   has no colour

red   green   blue   yellow . . .
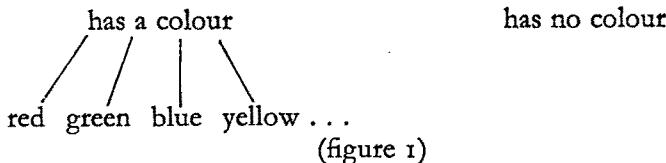
(figure 1)

On this view, since 'has a colour' is *false* of the number 2, then '2 is green', '2 is blue', etc. are all without truth-value. Without attempting to make out a case for this semantic, presuppositional treatment of sortal incorrectness, let me note that the natural way of applying it to semantic concepts is as follows:

sentences

bivalent   non-bivalent
(with truth-value)   (without truth-value)

true   false
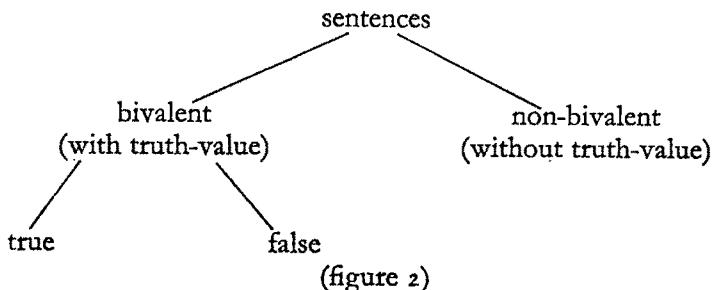
(figure 2)

And *this* arrangement yields M1—the sortal range of 'true' and 'false' consists of bivalent sentences. The relevant alternative, in which *sentences* constitute the sortal range of 'true' and 'false' (figure 3), is unnatural because it places 'non-bivalent' on a level with 'true' and 'false' as though it were a third truth-value:

sentences

true   false   non-bivalent
                                              (without truth-value)
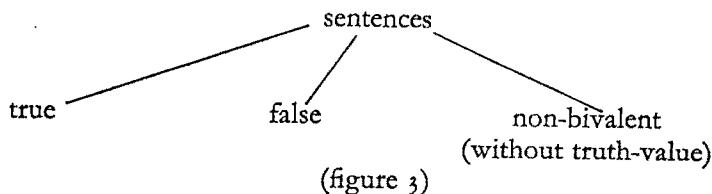
(figure 3)

So there is a connection between the view that sortally incorrect sentences are non-bivalent and the thesis M1, though the connection is not an equivalence. I had a clearer grasp of the argument in CSL (p. 92):

> The most likely alternative would be: RA (T) = RA (F) = set of sentences (i.e., with or without truth-value). The argument for this alternative would be something like the following. Every sentence is exactly one of the three: true, false, neither true nor false. If you say a sentence is true when in fact it is one of the other two possibilities, you have said something false. But this line of argumentation would break down

predication restrictions connected with any categories. For example, one might say: a given object is exactly one of the following: red, green, blue, yellow, . . . , none of these. So if I say of some object (say, virtue) that it is red, when in fact it is one of the other possibilities, then I have said something false. (This is not a defense of treating 'category mistakes' as without truth-value; it is only a defense, for one who already agrees to that, of restricting the RA (T) to the set of sentences with truth-values).

If one is unconvinced that the semantic, presuppositional treatment of sortal incorrectness is correct, one should count this feeling against the category solution when it comes time to compare that solution with others. From the beginning I have thought it a possibility worth exploring that the advantages apparently enjoyed by the category approach (e.g. accounting for the intuitive legitimacy of certain self-referential sentences with semantical predicates) outweigh such disadvantages as truth-value gaps for sortally incorrect sentences (if that is a disadvantage).

II

As for the inconsistency of M1 with M2–M5: strictly speaking, M5 is rejected by the category solution in its present form.[1] M5 is simply the statement:

(M5) 'Saturday flies' is not true or false;

I admit that there is ample reason for thinking that M5 is a statement sanctioned by the category approach. (Indeed, I *wrote* it in TSLP; but TSLP also contains a discussion (p. 307) which blocks Hugly and and Sayward's use of M5 to derive a contradiction—in particular, that discussion allows an exception to M3. More on this later.)

Strictly, it is M5':

(M5') 'Saturday flies' is not bivalent

(or, ' "Saturday flies" is without truth-value') that belongs to the category approach. M5' is consistent with M1–M4.

It is part of the category approach that the negations of truth-valueless sentences are also truth-valueless. Thus, looking back at figure 1, we see that not only '2 is green' but also '2 is not green' is without truth-value; similarly, not only ' "Saturday flies" is true' and ' "Saturday flies" is false' but also ' "Saturday flies" is not true' and ' "Saturday flies" is not false' are without truth-value.

Now on any reasonable view of the connectives, it follows that:

'Saturday flies' is neither true nor false

---

[1] The best published guide to the present form of the category approach is ORT.

(i.e., ' "Saturday flies" is not true and "Saturday flies" is not false')
is non-bivalent, since both component sentences are non-bivalent.

This is the view I should have taken from the beginning, especially
since, in my first paper (TSLP, n. 29) I adopted a position on compound
predicates with this view as a consequence. But at that stage I felt the
need to make an exception: 'I propose that we understand a disjunctive
predicate that exhausts a whole category as having the range of applic-
ability of the single predicate which is the family name of that category'.
(TSLP, p. 307) (This denies M3, the principle that makes the range of a
disjunctive predicate the intersection of the ranges of its disjuncts). I
said that the predicate 'true or false', because it exhausts the category
of bivalent sentences, *is* applicable to non-bivalent sentences. I had to
add, of course, that in general the sortal correctness of a sentence of the
form 'x is P or Q' does not entail the sortal correctness of 'x is P'.

To put the matter in the terms used by Hugly and Sayward: though
rejected by the category solution in its present form, M5 *does* hold in
TSLP. But M3 holds there only in a qualified form that allows for
exceptions for predicates such as 'true or false', so Hugly and Sayward's
argument to contradiction does not go through even for TSLP.

By the time I wrote SRCP and ORT I had fixed on the treatment of
the connectives that I now favour, and espoused M3 in its general
form. At this point I should have explicitly disavowed the exception,
but I did not.

My present view is that 'bivalent' is a semantical primitive, *not*
definable as, and not equivalent to, 'true or false'.[1] On the category
view, when a sentence is not bivalent one says precisely *that,* and
refrains from trying to apply 'lower' predicates whose use presupposes
that the sentence is bivalent. In general, the relationship shown in a
tree diagram between a category predicate and its family-member
predicates will not be expressible as an equivalence. I believe the
intuitive pull towards such statements of equivalence is reflected,
within the category view, by the distribution of truth-values and
truth-value gaps: *all* the family-member predicates are bivalent of an
object when (and only when) the category predicate is true of it.

I admit to forgetting to enforce this view: there are a few places
in my recent papers where I have written 'neither true nor false' instead
of, what I should have written, 'not bivalent'. But the intended 'official
position' is easy to infer—and, surely, once one notes the incom-
patibility of M3 and M5, as Hugly and Sayward do, then it should be
plain that M5 is simply to be replaced by M5'.

---

[1] This position was taken by John Pollock in 'The Truth about Truth' in Martin, ed.,
The Paradox of the Liar (New Haven, 1970).

III

Thus, in response to Hugly and Sayward's third criticism, I suggest that M5, which I once held, may easily be given up in favour of M5'.

California Institute of the Arts   © ROBERT L. MARTIN 1979

REFERENCES

TSLP Martin, R. L., 'Toward a Solution to the Liar Paradox', *Philosophical Review* 76 (1967), pp. 279–311.
CSL ——, 'A Category Solution to the Liar' in Martin, ed., *The Paradox of the Liar* (New Haven, 1970).
SRCP ——, 'Sortal Ranges for Complex Predicates', *Journal of Philosophical Logic* 3 (1974), pp. 159–167.
ORT ——, and Woodruff, F. W., 'On Representing "True-in-L" in L', *Philosophia* 5 (1975), reprinted in Kasher, A., ed., *Language in Focus: Foundations, Methods and Systems* (Dordrecht-Holland, 1976).

RIGID DESIGNATION

By H. W. NOONAN

IN Part I of this paper I propose an explanation of the facts about proper names which Saul Kripke has summed up in the thesis that proper names are rigid designators, one which allows for the application to proper names of Frege's sense/reference distinction. In Part II I try to explain away some of Kripke's examples of necessary truths knowable only *a posteriori* and contingent truths knowable *a priori*.

I

My thesis is that the facts summarised by Kripke in the statement that proper names are rigid designators are explained by the existence of a convention according to which *one must not use a proper name to speak of the sense one associates with it.*[1] I shall therefore be defending the Description Theory of Proper Names, when construed as a theory of

---

[1] Perhaps one uses a name to refer to the sense one associates with it when one asserts or denies the existence of the bearer of the name. Perhaps also one may use a name to refer to the sense one associates with it if one uses it within the context of an operator such as 'It is true that . . .', which converts an expression standing for a thought into one standing for the corresponding truth-value—see Dummett, *Frege: Philosophy of Language*, Duckworth, 1973, p. 277. Both these possibilities refute my suggestion as stated, but neither has any bearing on what I *mean*. The statement in the text must therefore be taken as rough; stated more exactly what I am arguing is that a convention exists whereby one is debarred from using a proper name to speak of the sense one associates with it in a context where replacement by a proper name with the same ordinary reference but a different sense is not guaranteed to preserve truth-value.

speaker's reference. But I shall not be defending the Description Theory of Proper Names construed as a theory of the reference of a name in a community, or denying that there is a great difference in semantic function between proper names and definite descriptions.[1] In fact, if this convention exists and has the basis I shall suggest, it is an im-: mediate consequence that proper names and definite descriptions are semantically very unlike. (It is important to see the difference between my suggestion and the fairly popular view that the sole difference between proper names and definite descriptions is the existence of a convention according to which proper names, but not definite des- criptions, must be assigned a maximal *scope* in any sentences in which they occur. This view could only be right if the Description Theory of Proper Names was a correct theory, not merely of speaker's reference, but also of the reference of a name in a community[2].)

My suggestion, to repeat, is that (with the qualifications in the foot- note on p. 174 there is a convention in force whereby one must not use a proper name to refer to the sense one associates with it, but only to refer to its bearer or, in quotation, to itself. Frege held both that proper names *could* be used in all three ways and that they *were* used in all three ways. The present suggestion agrees with the first half of this claim—for what we *cannot* do we do not refrain from doing by con- vention—but disagrees with its second half.

I shall try to make the suggestion plausible by answering three objections that could be made to it:

(1) First of all, the suggestion cannot be maintained unless Kripke's putative counter-examples to the Description Theory of Proper Names as a theory of speaker's reference can be dealt with. For to hold that it is only by convention that one is debarred from using proper names non-rigidly, or to refer, within modal contexts, to the senses one associates with them, one must maintain that on each occasion of use they are associated by the name user with a sense which fixes the referent on that occasion. Otherwise the *possible* ambiguity of utterances of modal sentences containing proper names (i.e. the one excluded only by the convention) which one is committed to in disagreeing with Kripke becomes inexplicable.

(2) Secondly, seriously to maintain the existence of a convention whereby one must not use a name to refer to the sense one associates with it one must explain the rationale of this convention.

(3) Thirdly, there are facts which it is very tempting to explain by appeal to the possibility of using a name to refer to the sense one

---

[1] For the distinction between a theory of speaker's reference and a theory of the reference of a name in a community see Gareth Evans' paper 'The Causal Theory of Names' in the *Proceedings of the Aristotelian Society*, supplementary volume XLVII, 1973.

[2] I am grateful to a referee for ANALYSIS for bringing this difference home to me and for several other helpful comments.

associates with it. For example, 'Ralph believes that Tully denounced Catiline' would seem to be equivalent to 'Ralph believes that Cicero denounced Catiline' with all the proper names understood as standing for their bearers, but on an ordinary understanding these sentences are not equivalent. Again, 'Ralph believes of Tully that he denounced Catiline' might be true when 'Ralph believes that Tully denounced Catiline' is false. If my suggestion is to be upheld these facts must be given another explanation.

Of these objections the third is evidently the most immediate and also, I think, the easiest to answer. I shall use some terminology and notation from a recent paper of Geach's in doing so.[1] I shall use the term 'aspect' to mean the Fregean sense of an (actual or possible) proper name, and speak of an aspect α as an 'aspect of' an object x when, in Fregean terminology, α is a mode of presentation of x, a way of latching on to x in thought. However, as well as aspects which are not embodied in the use of any actual proper name, there are aspects which are not the aspects *of* anything, as there are actual, but bearerless, proper names. Finally I shall write such things as '[α can F]'. This puts together a dodge belonging to Quine's theory of quasi-quotes with his use of square brackets to put a ring-fence around intentional contexts. In '[α can F]' the Greek letter does not belong to the intentional context: the thought that [α can F] is the thought you would express in language by attaching the predicate '—— can F' to a subject whose sense is the aspect α.

Now I suggest that 'Ralph believes that Cicero denounced Catiline' has the truth-condition:

(i) for some α, 'Cicero' expresses α in my or Ralph's idiolect and Ralph believes that [α denounced Catiline]

(here of course quantification is over aspects; I have ignored the occurrence of 'Catiline' for brevity).[2]

'Ralph believes that Tully denounced Catiline' has the truth-condition:

(ii) for some α, 'Tully' expresses α in my or Ralph's idiolect and Ralph believes that [α denounced Catiline].

Since (i) and (ii) are not equivalent nor are the sentences they

---

[1] 'Two kinds of Intentionality', *The Monist*, 1976, see especially p. 136. Cf. Kaplan 'Quantifying In' in Davidson and Hintikka eds. *Words and Objections*, Reidel, 1969, especially Section IV.

[2] Perhaps one may also say that Ralph believes that Cicero denounced Catiline if Ralph believes the same thought as someone who believes that Cicero denounced Catiline, i.e. where 'p' ranges over thoughts, if for some p, for some x, for some α, Ralph believes that [p], 'Cicero' expresses α in my or x's idiolect, x believes that [α denounced Catiline] and the thought that [p] is the thought that [α denounced Catiline]. But I ignore this complication in what follows.

analyse. But to explain this non-equivalence we have not had to interpret these sentences as ones in which the names 'Cicero' and 'Tully' are used to refer to senses. They are non-equivalent because the first contains a mention of the name 'Cicero' where the second contains a mention of the name 'Tully', and the first ascribes to Ralph a belief he could express by using 'Cicero' or some translation while the second ascribes to Ralph a belief he could express by using 'Tully' or some translation.

Again, the non-equivalence of 'Ralph believes that Tully denounced Catiline' and 'Ralph believes of Tully that he denounced Catiline' can be explained as follows. The former has the truth-condition (ii), the latter the truth-condition:

(iii) for some α, α is an aspect of Tully and Ralph believes that [α denounced Catiline].

and (iii) does not entail (ii) because (iii) could be true even if Ralph had never heard the name 'Tully' and had *no* name in his repertoire of which it was a translation. Nor does (ii) entail (iii) since (ii) could be true even if (iii) was false if there was no such person as Tully.

It seems to me that this line of reply to the third objection is a plausible one. The idea that the difference between someone who only believes *of* a person N that he is so-and-so and someone who believes *that* a person N is so-and-so is a difference between someone who could not, and someone who could, express his belief using the name 'N' or some translation is an idea that many have found tempting independently of the answer it provides to this objection. The same holds for the view that 'Ralph believes that Tully . . .' and 'Ralph believes that Cicero . . .' differ by *mentioning* different names and ascribing different linguistic capacities to Ralph.

Moreover this response to the third objection, i.e. the assignment of the truth-conditions (i), (ii) and (iii) to the relevant sentences, fits in well with Frege's idea that the sense of an ordinary proper name is different from speaker to speaker. This might seem hardly a recommendation of it. But a version of the Fregean idea which is at least plausible and I think in fact correct is that our use of ordinary proper names exemplifies what Putnam has called 'division of linguistic labour'. In other words it is characteristic of the use of an ordinary proper name that some of its users are deferential in their use of it to others and have no identifying knowledge of the name's bearer which can be expressed without reference being made to those others and to the sense that *they* attach to the name. That is to say, in a typical community of name users $X_1 \ldots X_n$ some of $X_1 \ldots X_n$ will attach to a name 'N' such senses as 'what X_3 calls "N"', 'what most of X_1 to X_{27} call

"N" '.[1] Of course, this will not be so for all of X_1 to X_n if the name in fact has a bearer, but it will be true for some and known implicitly by all to be true for some. And so the name will be, and will be known to be, in a sense ambiguous. However, the ambiguity will not be of a straightforward kind; for the users of 'N' will not be referring *by accident* to the same object. Rather, in associating such a sense as 'what X_3 calls "N" ' with a name one guarantees that whatever X_3 refers to by 'N' one will also refer to oneself. Thus in a community of name users, some of whom are deferential towards others, it will not be accidental, given the various senses they associate with the name, that they are all using it to refer to the same thing. If it *is* accidental for two name users then this can be traced back to two non-deferential name users who are accidentally referring to the same thing despite associating different (non-deferential) senses with the name. And then we can divide our linguistic community into two, grouping people according to which of these two non-deferential name users they are (directly or indirectly) deferential to (of course, the answer in some cases may be both, and then it is chance that reference is achieved at all). This suggests a criterion for joint membership in a community all of whom in an intuitive sense use a given name in the same way. Two people are joint members of such a community with respect to their use of a name 'N' if and only if either they attach the same sense to the name 'N', or one of them is deferential to the other in his use of 'N', or one of them is deferential to a third person who is deferential to a fourth . . . who is deferential to the other, or both of them are deferential to some third person—either directly or in the indirect way just indicated.

This is all very sketchy, but I think it suffices for the points I want to make. These are, first, that given that *some* such account of the typical functioning of proper names is correct, the view that the difference between 'Ralph believes *of* Tully . . .' and 'Ralph believes *that* Tully . . .' is parallel to that between 'Ralph believes *of* the man on the beach . . .' and 'Ralph believes *that* the man on the beach . . .', which is entailed by the attempt to explain the former difference as due to 'Tully' having a reference to its own sense in the second sentence, cannot be maintained— for if Ralph asserts 'the man on the beach . . .' I know what the content of his belief is, but if he asserts 'Tully is . . .' I can only guess at the content of his belief and shall very likely guess wrong. Hence—and

---

[1] To complicate matters there is also the possibility of a sense of the type 'That which is denoted by most of the identifying descriptions associated by the members of this community with "N" or, if there is no such object, the such-and-such'—where 'the such-and-such' is a non-deferential identifying description. The possibility that all the members of a community should associate senses of this type with a name seems to be needed to make intelligible Dummett's example of a name of a seldom visited wood, where each user of the name has a different criterion of identification available but intends in using the name to subordinate his private 'sense' to the majority reference. See *Frege: Philosophy of Language*, pp. 140-141.

this is my second point—some other explanation of the difference in
question is needed; and one which makes both of 'Ralph believes *of*
Tully . . .' and 'Ralph believes *that* Tully . . .' involve only quantification
over, and no identifying reference to, senses, as interpretation via
(iii) and (ii) does, seems appropriate.

   This brings us to the second objection, which can also be answered
fairly easily once difference of sense from speaker to speaker is acknow-
ledged as characteristic of our use of proper names—indeed the answer
is implicit in the foregoing. For we aim in conversing with people to
speak about the same things as they do—if we fail to do this, if we lack
a common subject-matter, we fail to communicate at all (as do two
people who agree that 'the bank is nearby'—one meaning the river
bank and the other the financial bank). But if one understands a sentence
containing a proper name as about the sense one associates with that
name, and everyone else does the same, then if sense varies from speaker
to speaker there *will* be no common subject-matter spoken of using
that sentence. Thus if we wish what we say to have the possibility of
agreeing or disagreeing with what others say, we must not understand
sentences containing proper names as about the senses *we* associate
with those proper names; we must either forego talk of senses
altogether or talk of them only in the manner indicated in the
reply to the third objection. But then the rationale of a convention
according to which one must not use a proper name to speak of the
sense one associates with it is clear: to flout this convention is deliber-
ately to make what one says irrelevant to what others say, as when
one man says 'The bank is nearby' meaning the river bank, and another,
knowing this is what is meant, replies, 'No, it isn't', meaning the financial
bank.

   Finally I come to the first objection. I shall be brief about this.
While there is doubtless a great deal more to say the essential point
seems to me to be this. Kripke says that one can refer to a thing by a
name, know and assert things about it, when one knows of no descrip-
tion which determines the referent of that name, not even one which
involves a first-person reference to oneself. But none of his examples
shows anything so extreme. It is true that someone, say Jones, may
use 'Feynman' to refer to Feynman knowing little more than that he is
a physicist. But this will not be quite all Jones knows. He will also
know that there is just one person such that (a) he is named 'Feynman'
and (b) his (Jones') familiarity with the name 'Feynman', as a name of a
physicist, derives from having encountered references to that person
by that name. Suppose Jones believes this to be so but it is not so:
there are two physicists named 'Feynman' and Jones has encountered
references to both of them made by that name without realising that
there are two. Then evidently enough Jones is not referring to anyone

when he uses the name 'Feynman'—he is merely in a muddle. But if Jones does know that there is just one such person, he can identify Feynman by the description 'the person such that (a) he is named "Feynman" and (b) my familiarity with the name "Feynman", as the name of a physicist, derives from having encountered references made to him by that name'.[1]

Again, while one does not use 'Gödel' to refer to Schmidt just because Schmidt proved the Incompleteness Theorem and the only biographical detail one associates with the name 'Gödel' is 'prover of the Incompleteness Theorem' this proves nothing as strong as Kripke claims. For to use 'Gödel' to refer to Gödel in this circumstance one must know that there is just one person such that (a) he is named 'Gödel' and (b) one's familiarity with the name 'Gödel', as the name of a logician, derives from having encountered references made to him by that name. But if one knows this one can identify Gödel by the description 'the person called "Gödel" such that my familiarity with the name "Gödel", as the name of a logician, derives from having encountered references made to him by that name.'

In general some such egocentric description is bound to be available, if nothing else is, whenever one is capable of referring to an item by a proper name, and of course such descriptions are *not* circular in the manner of 'the person I (here and now) mean by "Jones" '—they are genuinely reference-determining.

Moreover, quite apart from these particular considerations, there is a difficulty in principle with the idea of counter-examples to the Description Theory of Proper Names, construed as a theory of speaker's reference, which to my mind shows clearly that such counter-examples *must* be impossible. For if a man *intends* the reference of a name he utters to be the so-and-so then it must *be* the so-and-so—or nothing (of course, he can *believe* the reference of the name to be the so-and-so when it is not). If a man really intends the reference of 'Gödel' when he utters it to be the man who proved the Incompleteness of Arithmetic then that is who it will be—even if that is Schmidt. Thus when a speaker has intentions capable of determining a specific object as the reference of a name utterance, the Description Theory is bound to apply. To find counter-examples to it, then, one must apparently consider cases

---

[1] This description may denote even if 'the famous physicist named "Feynman" ' does not, for example if there are two famous physicists called 'Feynman' but by chance Jones has only ever encountered references made to one of them. But if Jones is a layman this situation is an unlikely one: more probably, if 'the famous physicist called "Feynman" ' does not denote nor will the description in the text, as used by Jones. This explains why Jones would be embarrassed by the information that there were two physicists called 'Feynman', both famous. If there were two physicists called 'Feynman', only one of whom was famous, however, then more than likely Jones, as a layman, would only have encountered references to the famous one, so that the description in the text, used by him, would still denote. This explains why *this* piece of information would not put Jones out of countenance.

where *no* specific reference-determining intention exists.[1] But however
the reference is determined in such a case it must be in a manner which,
if it were described to them, the speaker and the other members of his
linguistic community could be brought to acknowledge as correct, i.e.
as bound to determine the true reference on any relevant occasion of
utterance. Thus it must be a mode of reference determination implicitly
known to the speaker and *intended* by him to determine his reference—for
he knows that it *will* determine his reference in the absence of a specific
reference-determining intention on his part, and so lacking such an
intention he must *intend* to let it do so. And so once more the Description
Theory of Proper Names, construed as a theory of speaker's reference,
applies. The *most* that could possibly be the case is that sometimes the
identifying knowledge a speaker has of his reference on a given occasion
of utterance of a name is inexpressible except by way of a reference to
the very name *token* uttered on that occasion. But even if this were the
case (though there is no reason to suppose that it ever is) it would not
refute the Description Theory of Proper Names.[2]

I hope then to have shown in this part of the paper that all three
objections listed fail to refute my suggestion that there is a convention
according to which one must not use proper names to speak of the
senses one associates with them, and in doing so to have given some
reasons for looking on it with favour.

II

. In the light of the foregoing one can explain away as follows certain
of Kripke's examples of necessary truths knowable only *a posteriori*
and contingent truths knowable *a priori*:

(1) Necessary *a posteriori* truths: Kripke cites as some examples
identity statements in which the sign of identity is flanked by two proper
names, e.g. 'Cicero is Tully'. It follows from what I have suggested
that in most people's mouths this will express a *contingently* true thought
(different thoughts in different mouths of course). The reason for the
appearance of necessity is that when the operator 'Necessarily' is
prefixed to this statement—yielding 'Necessarily, Cicero is Tully'—the
resultant statement is one to which the convention that one must not

---

[1] This inference, as the referee pointed out, depends upon the thesis that an intention
capable of determining the reference of a name utterance must be of the type: to refer by
'N' to the φ—where 'the φ' may of course contain an indexical component. To deny this
one must hold that the reference of a name utterance may be determined by an intention of
the type: to refer by 'N' to M—where 'M' is itself a proper name. But this is
incomprehensible, as becomes immediately evident when one thinks of the case in which the
speaker knows *several* people called 'M'. A case in which a name user succeeds in referring
despite *all* his intentions with regard to his reference being of this latter type can only be a
case in which something *other* than his intentions is determinative of his reference.

[2] Not the one Kripke states in 'Naming and Necessity' (in *Semantics of Natural Languages*,
eds. Harman and Davidson, 1972) anyway, which is the only one it is at all plausible to
suppose Frege was committed to.

use a proper name to refer to the sense one associates with it applies. Hence 'Necessarily, Cicero is Tully' must be understood in accord with this convention, as equivalent to 'Concerning Cicero and Tully: necessarily the former is identical with the latter', and since identical objects are necessarily identical this is true. But there is no *thought* here which is both necessarily true and knowable only *a posteriori*.

(2) Contingent *a priori* truths: Kripke gives as his example 'The length of this rod is one metre', where 'this rod' is the standard metre rod. Another example acceptable to him would be Dummett's 'St.Anne was a parent'. I shall apply what I have to say to this example. In the mouth of someone who means by 'St. Anne' 'the mother of the Blessed Virgin Mary' this statement expresses a *necessarily* true thought. The reason for the appearance of contingency is that when the operator 'Necessarily' is prefixed—yielding 'Necessarily, St. Anne was a parent'— the resultant statement is one to which the convention that one must not use a proper name to refer to the sense one associates with it applies. Hence 'Necessarily, St. Anne was a parent' must be understood in accord with this convention, as equivalent to 'Concerning St. Anne: necessarily she was a parent', and since parents are not necessarily parents this is false. But there is no *thought* here which is both contingently true and knowable *a priori*.

Of course, there are other candidates for the status of necessary *a posteriori* and contingent *a priori* truths than those here considered. For example, 'Every man has such and such a DNA structure' (necessary *a posteriori*) or 'Everything actually red is red' (contingent *a priori*). I think that some of these can be dealt with by an extension of my method, together with an appeal to our intuitive classification of properties, not only of objects but also of concepts, as necessary or contingent. For others other treatments seem appropriate. For example, the claim that 'Everything actually red is red' is *a priori* but non-necessary seems quite parallel to, and just as spurious as, the claim that 'Everything presently red is red' is *a priori* but non-eternal. But, however this may be, it seemed worthwhile to show how the (extremely popular) candidates for necessary *a posteriori* and contingent *a priori* status considered above could be dealt with.

LOCKWOOD AND MILL ON CONNOTATION AND PREDICATION

By B. F. KEATING

I

IN 'On Predicating Proper Names', Michael Lockwood attempts to solve a problem in J. S. Mill's semantics.[1] On the one hand, Mill holds that names do not connote. But, on the other hand, he holds that names may serve to predicate. Just as one may predicate the attribute of being a man by saying, 'That is a man', one may, Mill suggests, predicate something by saying, 'That is Jimmy Carter'. However,

> . . . if a predicate is an expression which is used to predicate attributes, and if the attributes which it serves to predicate are those which it connotes, Mill is faced with a problem. For how, then, can he consistently maintain that proper names can function as predicates, when he denies that they connote anything at all? (p. 489).

Lockwood's answer may be summarized as follows: the spirit of Mill's claim that names lack connotation is that 'their standard function in utterances is referential, rather than attributive' (p. 496). Where $S(t)$ is a sentence that contains a definite singular term t and that is used to make an assertion, t functions referentially only if 'there is an object X such that

(i)   it is a necessary and sufficient condition of the truth of what $S(t)$ is being used to assert that X satisfy the predicate expressed by the context $S(\ )$; . . .' (p. 485).

But this 'Kripke-Donnellan conception' of names as purely referential devices is compatible with

(A)   . . . a proper name in predicate position will serve to ascribe what it connotes to the individual denoted by the subject term (p. 496),

if the predicating name is regarded as connoting the attribute of being (identical to) X. Hence the spirit, if not the letter, of Mill's theory is not subject to the above problem about predication (p. 496).

In Section II I attempt to clarify Lockwood's view of connotation, and in Section III I object to the thesis that connotative names can function referentially. Finally, in Section IV, I present evidence that Mill would reject Lockwood's view of 'propositional content'.

II

Lockwood claims that names in predicate position will predicate what they connote, but he seems to interpret 'connote' in a number of ways. Both Mill and Lockwood regard connoting as a relation between a term and an attribute (ignoring Mill's phenomenalism). Where P in an utterance $\ulcorner S$ is $P\urcorner$ is a referentially functioning proper name or an attributively functioning definite description, Lockwood apparently holds that P connotes an attribute if and only if that attribute figures in the propositional content of the utterance. Thus:

> [the distinction] between what the term connotes (in the given context) and what, by way of its connotation, it actually contributes to the propositional content of the utterance in which it figures . . . is idle in the case of proper names, and connotative terms that are functioning attributively (p. 496).

But how do attributes connoted by a predicate term. 'figure' in propositional content? Lockwood does not explicitly say but the following description of attributive functioning is relevant:

> Upon discovering, to my regret, that an attractive girl whom I have just met is married, I might remark, 'Her husband is a lucky man'. Here I am saying that whoever has the *attribute* of being that girl's husband also has the attribute of being lucky. What is connoted by the words 'her husband' enters essentially into the proposition being expressed (p. 483).

Though this is about subject terms it suggests

> (B) If a definite singular term P in $\ulcorner S$ is $P\urcorner$ connotes an attribute, and S functions attributively, then the proposition expressed may be represented as *that S is the F,* where what P connotes is the attribute of being the F.

Our passage suggests that if P is 'her husband' (used attributively) then being the F is being the one who is that girl's husband. Where P is, say, 'Jimmy Carter', and 'Jimmy Carter' denotes X, Lockwood elsewhere implies that being the F is being the thing identical to X: a name connotes 'being identical with a certain individual' (p. 474).

(B) is compatible with Mill's account of propositional content for connotative predicates: one's assertion, he says, is to the effect that such and such 'has the attributes connoted by the predicate' (quoted in Lockwood, p. 491). Also, where one's subject term functions attributively, (B), plus a Millian thesis, yields Lockwood's (A). Surely a proper name will not serve to ascribe attributes unless it has a meaning. But Mill says that the only terms which 'have properly speaking any meaning' are such that 'the meaning resides not in what they *denote,* but in what they *connote*' (quoted in Lockwood, p. 493). So a predicating name will connote. But then what is asserted will be to the effect that

such and such is the F, where the F is what is connoted. Therefore, the name will serve to ascribe what it connotes to the denotation of the subject term.

The claim that meaning resides in connotation did not lead Mill to abandon his thesis that names lack connotation. It probably does follow from an equation of a *term's* connotation with its contribution to propositional content that names connote, but Lockwood rejects such an equation. On his view, a definite description may both connote and function referentially, and yet *not* predicate its connotation. Such a term, though it will then be 'serving merely as a surrogate for a proper name' (p. 484), will not contribute its connotation to propositional content (p. 496). This makes it difficult to see why Lockwood's equation of a name's connotation with its contribution to propositional content should hold.

Regarding the 'Her husband is a lucky man' case described above, Lockwood says,

> . . . if any individual is to stand to 'her husband' in the relation that clause (i) requires, . . . it will be the man who is actually married to the girl under discussion (p. 485).

If the individual does not fit the description used, then there will not be the 'convergence between words and intention' necessary for 'genuine assertion' (p. 486n). So the attribute which is connoted when the description functions attributively—namely, that of being her husband—must, assuming genuine assertion, be satisfied by the referent in the referential case also. Now, in Lockwood's terminology, a 'connotative' term may function either attributively or referentially. If such a term maintains the same connotation in either case, then, assuming genuine assertion, the attributes connoted in the referential case will be those one may 'read off', so to speak, from the description used. This may explain Lockwood's view that one may predicate attributes distinct from those connoted. In saying 'Here comes her husband', I may be generally known to mean merely: here comes the fellow who acts as though he is her husband. Moreover, I may say what I mean and speak truly when it is not really her husband. In this case, the attributes I predicate are not those connoted. Of course Lockwood must also say, implausibly, that I do not make a genuine assertion.

However, on this view of connotation, (A) need not be true. For I may similarly speak truly saying of an approaching phenomenology enthusiast 'Here comes Husserl himself', it being clear that I mean that fellow who writes just like Husserl. In this case the attributes I ascribe will not be those connoted, on the 'read off' view of connotation. Let us set aside this interpretation of connotation in favour of (B). (B) captures Lockwood's equation of a name's connotation with its contribution to propositional content.

III

Suppose that 'Jimmy Carter' serves to predicate in 'The man in the corner is Jimmy Carter.' By (A) the name connotes. By (B) the propositional content is that the man in the corner is the F (assuming the description functions attributively). Now suppose the name functions referentially. It follows that there is an object X such that a sufficient condition of the truth of what the sentence is used to assert is that X satisfies 'The man in the corner is ——'. But this is false. X must also be the F if what is asserted is to be true. Lockwood will reply that, in this case, being the F *is* being identical to X. For example, if X is Jimmy Carter then the connotation is being (identical to) Jimmy Carter. But in order for X to *be Jimmy Carter*, X must be historically related to 'Jimmy Carter'. So it is not sufficient for truth that X satisfy 'The man in the corner is ——'. X must also be historically related to 'Jimmy Carter'. So if the name connotes, it does not function referentially in Lockwood's sense.

Of course it *is* a necessary condition of the truth of any assertion about an item X that X is self-identical. An object X will 'have to' satisfy that attribute in addition to satisfying whatever attributes figure in the propositional content of the utterance in order for what is asserted to be true. Similarly, X will 'have to' satisfy the attribute of being such that $2 + 2 = 4$. But not every necessary condition of the truth of what one says can be a part of what one says, and it is the nature of what is said that is in question in giving connotation or contribution to propositional content. Being the thing identical to X cannot be interpreted as being self-identical, for the latter attribute is clearly not what I predicate of the man in the corner when I say (perhaps falsely) that he 'is Jimmy Carter'.

The proposition is standardly equated with what is asserted. What I use the above sentence to assert may be my belief that the man in the corner is the one commonly referred to as 'Jimmy Carter'. If the attribute associated with 'Jimmy Carter' by way of giving its contribution to the propositional content of my utterance is the attribute connoted by the name, the attribute connoted will be 'meta-linguistic' in character. To this Lockwood will reply:

> For an utterance to be true its words must no doubt stand in certain characteristic relations to the world and to one another; but to suggest that in using these words to make a statement we are thereby asserting that these relations hold is to imply that we can never use words without also mentioning them (p. 490).

Though I agree with this as far as it goes, I want to point out that at least some of the very words one uses in making a statement might correctly be *mentioned* in an account of *what one says*. Jones, who is ignorant of mathematics, reads in a maths book: 'Fuchsian functions are

not commonly understood' and believes what he reads. It may be that all he can be credited with saying, when he later assertively utters this sentence at a party, is that what mathematicians call 'Fuchsian functions' are not commonly understood. If so, what he says is in a proper sense of 'about' about the words 'Fuchsian functions', which he uses. Though these words are not mentioned in expressing what is said, they *are* mentioned in an account of what is said. (Here I am indebted to James Cargile).

The general attempt to exclude 'meta-linguistic' matters from propositional content raises difficulties. In saying 'A bachelor is an unmarried male', one implies that 'bachelor' and 'unmarried male' coapply. If this implication is due only to 'pragmatics' and not to propositional content (see p. 490), then it will be difficult to distinguish the *content* of that utterance from the content of 'A bachelor is a bachelor'.

Lockwood writes

> . . . for 'her husband' to be serving a purely referential function [in 'Her husband is a lucky man'], the truth conditions of the statement that [the sentence] is being used to express must be: That Jacob Jones is a lucky man (p. 486),

where Jacob Jones happens to be the referent. But it is dubious whether, in using 'her husband' referentially simply to indicate a certain person, one expresses a proposition that entails that Jacob Jones exists. If I am right, such a proposition implies the existence of someone historically related to 'Jacob Jones', and this is certainly not an implication of what I say using the description. But then, what proposition do I express? We have already noted that, on Lockwood's view, the attribute of being her husband does not figure in the proposition, even though something must be her husband to be the subject of the 'genuine assertion'. Without an account of what is said we are not in a position to give the term's connotation or contribution to what is said.

IV

In uttering $S(t)$ I may predicate $S(\ )$ of an item X. And, it *is* a necessary and sufficient condition of the truth of that predication that X satisfies $S(\ )$. Lockwood characterises the 'significance' of a proper name as its 'contribution to propositional content' (p. 493). If 'propositional' is read as 'predicational', Lockwood's further claim that 'a name has no significance over and above its denoting the object that it does' seems to follow.

But there is some evidence that Mill would not have accepted an equation of the proposition with the predication. He writes in his chapter on the 'import of propositions':

> When we say Socrates was contemporary with the Peloponnesian War, the foundation of this assertion, as with all assertions concerning substances, is an assertion concerning the phenomena which they exhibit . . . although the proposition does not assert that alone.[1]

This implies that the proposition does assert at least that, but what is 'that'? Mill suggests that Socrates himself is not a constituent of the assertion; he is 'to us nothing but that which causes . . . phenomena'. If so, then the assertion is not simply the predication of Socrates, however he is 'exhibited', of the attribute of being contemporary, etc. It seems more reasonable to say that the assertion is that the cause of certain phenomena is contemporary, etc.

Mill elsewhere gives a general account of propositional import for non-connotative subjects that runs thus: 'the individual thing denoted by the subject has the attributes connoted by the predicate term' (see p. 491). This suggests that 'Socrates', in subject position, should not be regarded as connoting an attribute. Mill's reason for suggesting this seems to be that, in a sentence such as the above sentence about the Peloponnesian War, 'Socrates' is not applied to Socrates to *signify* that he possesses any attribute. But here a different notion of connotation is being employed. For this fact about the application of 'Socrates' does not imply that the name should not be associated with an attribute by way of giving its contribution to propositional content. Indeed, when speaking of the import of propositions, Mill does not equate the proposition expressed with the predication. To this extent, he disagrees with the referentialist account of how names contribute to propositional content.

West Virginia Wesleyan College   © B. F. Keating 1979

---

[1] J. S. Mill, *A System of Logic*, George Routledge and Sons, Ltd., (London, 1891) p. 72.

A QUESTION OF CONNOTATION: AN ANSWER TO KEATING

By MICHAEL LOCKWOOD

KEATING'S article (ANALYSIS 39.4, October 1979) provides a welcome opportunity for me to consider anew some of the issues raised by Mill's treatment of proper names. The 'problem in Mill's semantics', to which, as he says, my earlier paper was largely addressed, arises out of the following three claims, each of which finds explicit expression at some point in *A System of Logic*:

  (a) When a term functions as a predicate, the attributes it connotes are thereby 'predicated of' the subject of the sentence in which the term figures;

  (b) proper names can function as genuine predicates;

but (c) proper names connote nothing.

Keating does not deny either that Mill *is* committed to (a), (b) and (c), or that these assertions form an inconsistent triad. Something has to give here, and what I attempt to establish in 'On Predicating Proper Names' (*Philosophical Review*, LXXXIV, 1975, pp. 471–498) is that, contrary to (c), proper names do, after all, connote. Keating, however, is puzzled by my reasoning. First, he suggests that I 'interpret "connote" in a number of ways', inasmuch as, following Donnellan, I hold that the connotation of a definite description sometimes enters into the 'propositional content' (what gets said) by an utterance in which the name appears, and sometimes does not, merely serving to fix a referent for the utterance. The plain implication, here, is that my use of the expression 'connote' is somehow ambiguous. I find this charge obscure. Would Keating say, I wonder, that someone was interpreting 'knife' in a number of ways, were he to maintain that a knife (of the kind that commonly accompanies a fork) may standardly be used either to cut or to spread? But let it pass. Keating has a serious point when he observes, in effect, that my application, to definite descriptions in ostensible predicate position, of something analogous to Donnellan's referential-attributive distinction, already commits me to denying, in the general case, the equation implied in (a) between what is predicated by a term and what it connotes. Hence I cannot (and indeed do not) argue simply from (a) and (b) to not-(c). What I in fact argue, as Keating recognizes, is that (a) does hold for the special case of proper names. But Keating finds it 'difficult to see' why I should insist on this.

  The answer is that what is connoted by a definite description, occupying predicate position, yet functioning referentially, is, as I see it, the possession of a certain specific identity. Since, on my view,

genuine proper names invariably connote just that, possession of a certain identity and no more, and since what is predicated in an utterance whose predicate is an attributively functioning singular term is whatever the term connotes, there is simply no room, here, for a non-referential use of proper names (as proper names) in which they serve to predicate anything other than bare possession of the identity in question. It was on these grounds that I maintained, in my original article, that for proper names the very distinction between attributive and referential use lapses.

This, however, Keating seeks to challenge, in Section II of his paper, on the strength of what he sees as a close parallel between the following two cases:

Case I. I say 'Here comes her husband', being 'generally known to mean merely: here comes the fellow who acts as though he is her husband'. In this case, Keating argues, 'I may say what I mean and speak truly when it is not really her husband'. For here, he thinks, 'the attributes I predicate are not those connoted'.

Case II. I say of an approaching phenomenology enthusiast—once more, truly, according to Keating—'Here comes Husserl himself', 'it being clear that I mean that fellow who writes like Husserl'. Here again, says Keating, 'the attributes I ascribe will not be those connoted . . .'.

Let me say at once that I am not precisely sure just what these examples are intended to be examples of. Specifically, are we to understand them as involving supposedly referential uses of the key singular terms, or attributive ones? The preamble strongly suggests the former. But then, since referential uses are, both on Donnellan's construal and my own, purely 'entity-invoking' (in C. A. B. Peacocke's sense, 'Proper Names, Reference and Rigid Designation' in *Meaning, Reference and Necessity,* ed. S. Blackburn, Cambridge 1975), Keating cannot consistently say what he seems to say, namely that what are respectively predicated, in the two cases, are the attributes of acting like so-and-so's husband, and writing like Husserl.

This aside, what is the correct thing to say about these examples? Well first, I agree with Keating that there is a certain parallel between them. In short, both appear to exemplify *ironic* uses of language. Now I do not have a theory of irony, but I would have thought one would find a fair measure of agreement that to speak ironically is to say one thing and mean (i.e. imply) another. This does not square with Keating's claim, in regard to Case I, that 'I . . . say what I mean, and speak truly'. On the contrary, I would have thought that both were fairly clear cases of *not* saying what one meant, and not speaking truly. Has the speaker, in either case, made a genuine assertion (in the sense of having asserted

some proposition, not just, trivially, having performed a declarative speech act)? Keating thinks I am committed to saying not, in view of my insistence on the need for a 'convergence between words and intention'. But one must be careful here; a speaker may say what he means to say, without saying what he means. 'Intention', in the above phrase, is to be understood as cognate to 'intended propositional content', not 'what information one intends to convey'. In Case I, I am still inclined to say that no assertion has been made, if my utterance of the words 'her husband' is masquerading as a referential use of the phrase. But as far as I can see there is no bar, in Case II, to saying that I really do intend to assert (falsely) that Husserl is approaching, and succeed in so doing, meaning, however, to convey thereby the correct information that someone is approaching of whom it is common knowledge, as between my audience and myself, that he writes like Husserl. Parallel considerations would allow one, in Case I, to hold that I had deliberately said something false, were I using the phrase 'her husband' attributively, and were the woman in question in fact married (though not to the individual whose approach triggered my remark). The point is that in neither case, unlike those considered by Donnellan and myself, is the speaker labouring under any kind of misapprehension.

Thus, I cannot see that 'Husserl', in the second example, need actually be construed as being used to predicate any attribute other than that which, on my view, it connotes: that of being Husserl, if what is involved is irony, or possibly, if the appellation has come to attain the status of a nickname, having the identity of the phenomenology enthusiast so rechristened.

Suppose, however, I am wrong about this, and Keating right. What would follow? A theory of meaning has, after all, to start with the central cases and work out, not the other way around. From uses which are so obviously secondary ('parasitic', as an earlier generation of philosophers would have said, on the standard case), I do not see what general moral could possibly be drawn for the normal use of proper names in predicate position.

Nor, seemingly, does Keating, for he does not pursue this line of argument. Rather, he proceeds in Section III to argue, first that my account of the connotation of proper names yields the wrong truth-conditions for utterances in which they figure as grammatical predicates, and secondly that only some metalinguistic account (such as Mill himself advances) can provide the right truth-conditions here. The issue between us is this. It follows from my account that if someone utters a sentence of the form 'The man in the corner is ——', the term occupying the blank will be functioning referentially only if there is some individual whose satisfaction of (what is expressed by) the corresponding open sentence suffices for the truth of what is asserted. Where the sentence

uttered is 'The man in the corner is Jimmy Carter', there is, I maintain, such an individual, to wit Jimmy Carter. Keating, however, baulks at this. For the assertion expressed by the utterance to be true, Jimmy Carter is, he thinks, also required to be historically related in the appropriate way to (this use of) the name—thereby giving the lie to the suggestion that the name is, by my definition, functioning referentially. That, in fact, is what Jimmy Carter's *being* Jimmy Carter must come to, according to Keating, if the latter is to be what the name connotes, and the name's connotation is to enter appropriately into what gets said by utterances containing the name as predicate.

But there is a fundamental confusion here. In some suitable context, a White House reception say, I utter the words 'The man in the corner is Jimmy Carter', thereby making a certain statement, as it happens a true one. Clearly, my words would not have been capable of expressing the assertion that they do, had not Jimmy Carter the man stood to 'Jimmy Carter' the name in the requisite relation. But his so standing serves to contribute, here, only to my making a true statement, *not,* contrary to what Keating maintains, to the truth of what I state. *The statement I express* in uttering the words 'The man in the corner is Jimmy Carter' could have been true, even if Carter had never acquired the name he in fact bears; it is just that I should not then have been able to use these words to express that statement. By the same token, Jimmy Carter would still, on my theory, have *been* Jimmy Carter, even if he had never been so named. To be (identical with) Jimmy Carter is, as I see it, neither to be merely self-identical (for everything is self-identical, whereas only Jimmy Carter is identical with Jimmy Carter) nor, even in part, to be called 'Jimmy Carter'; it is to have the specific identity that Jimmy Carter, alone of all individuals, possesses, and this is what I take the name 'Jimmy Carter' to connote.

Keating thinks he finds an analogue for his metalinguistic view of the predicative function of proper names, in the assertive utterance by one Jones, ignorant of mathematics, of the sentence 'Fuchsian functions are not commonly understood', previously read in a maths book. Here, says Keating, what Jones is really saying is that what mathematicians call 'Fuchsian functions' are not commonly understood. Well, possibly. But is it not far more plausible to construe this as a *deferential* use, in Gareth Evans' sense ('The Causal Theory of Names', *Proceedings of the Aristotelian Society,* supplementary volume XLVII (1973), pp. 187–208) whereby Jones intends to assert (commit himself to the truth of) *whatever* statement his words would convey to a trained mathematician, without, however, knowing what statement that is. Note in passing that it will, on either account, be false to say of Jones, as does Keating, that on encountering the sentence he 'believes what he reads'. On the contrary, he is in no position to believe either what he

reads or (on a deferential construal) what he later says, for he does not, in the required sense, know what he is reading or saying; he merely believes that, whatever it may be, it is true.

In any event I cannot see that the model is, for Keating's purposes, a helpful or appropriate one. For, in the 'Fuchsian functions' case, and others of this ilk, there is invariably a primary use of the very same sentence which is neither metalinguistic nor deferential; and this Keating simply cannot allow in the case of such a sentence as 'The man in the corner is Jimmy Carter'. Thus, I agree with Keating that someone uttering this sentence *might,* in some sense, be expressing his 'belief that the man in the corner is the one commonly referred to as "Jimmy Carter" ' (though as we have seen it would not, as Keating thinks, follow that that was what was being asserted). But this would, in general, be an appropriate interpretation only where the speaker lacked, or took his audience to lack, a prior mastery of the name 'Jimmy Carter': where knowledge of who Jimmy Carter was, or knowledge of him under that name, was presumed to be absent. Any such metalinguistic account of what was expressed (let alone asserted), in the presence of such knowledge—actual in the speaker, presumed in the audience—strikes me as implausible in the extreme. For it is on a par with suggesting that a mathematician who knew what Fuchsian functions were, under that description, and took his audience to share that knowledge, would still, by way of such a sentence as 'Fuchsian functions are nonrecursive', succeed in expressing no more than his belief that whatever mathematicians call 'Fuchsian functions' are non-recursive—an analysis which both Keating and Mill would quite properly reject out of hand.

Finally, a word on Mill's 'phenomenalism'. Now Mill was, I dare say, a phenomenalist at the time of writing his *Examination of Sir William Hamilton's Philosophy* (London, 1865), where we find the celebrated dictum that matter consists of 'groups of permanent possibilities of sensation'. But *pace* Keating, what Mill espouses in *A System of Logic* (first published over twenty years earlier, in 1843) is not phenomenalism but causal realism (dressed up in the Kantian terminology then in vogue). Admittedly Mill was, even then, inclined to commit the error of supposing that the experiential grounds of a material-object assertion somehow constituted a part of its content. But, for all that, I am completely baffled by the following passage in Keating's article:

> Mill suggests that Socrates himself is not a constituent of the assertion [that Socrates was contemporary with the Peloponnesian War]; he is 'to us nothing but that which causes . . . phenomena'. If so, then the assertion is not simply the predication of Socrates, however he is 'exhibited', of the attribute of being contemporary, etc.. [Rather] the assertion is that the cause of certain phenomena is contemporary, etc.

What is Keating saying, here? Is he suggesting that Mill himself would have denied that 'Socrates was contemporary with the Peloponnesian War' expresses a *de re* assertion concerning Socrates, an assertion, that is, which says *of Socrates* that he was contemporary with the Peloponnesian War, and is accordingly true or false depending on the way things were with that individual? (I can give no other sense to Keating's metaphorical talk of Socrates' being 'a constituent of the assertion'). If this is the claim, then it is flatly contradicted by the text. In the very paragraph from which Keating quotes, Mill insists that the proposition that Socrates was contemporary with the Peloponnesian War

> asserts that the Thing in itself, the *noumenon* Socrates, was existing, and doing or experiencing ... the series of facts by which Socrates manifested himself to mankind, and the series of mental events which constituted his sentient existence, ... simultaneously with the series of facts known by the name of the Peloponnesian war (J. S. Mill, *A System of Logic*, 8th ed., Harper & Brothers, 1874, p. 82).

Perhaps, however, Keating deems it inconsistent of Mill to have held this, while maintaining that 'a substance [is] to us nothing but either that which causes, or that which is conscious of, phenomena'. But then, where *is* the inconsistency? Why should the fact, if it be a fact, that we are acquainted with Socrates only via the phenomena of which he is the cause prevent us from making *de re* assertions concerning him? Keating's own suggestion that from the standpoint of Mill's philosophy the term 'Socrates' should, under analysis, give way to 'the cause of such-and-such phenomena' totally overlooks the possibility that a speaker's grasp of the reference of the name 'Socrates' might be mediated by some such description, without the latter's forming, or or being taken to form, any part of the name's intrinsic significance.

In such relatively brief compass, it has proved impossible, inevitably, to pass comment on everything Keating has to say in his article. But I hope, at any rate, to have countered successfully those points which he would deem most crucial to his argument. My reconstruction of Mill's theory would seem, on the face of it, to have emerged unscathed from Keating's assault.

University College, Oxford   © MICHAEL LOCKWOOD 1979

ACTIONS AND IDENTITIES

By JENNIFER HORNSBY

IRVING Thalberg has recently proposed an account of the inviduation of actions. He contrasts it with two rival accounts, and he claims that its great merit is to reconcile the parties by showing that the most important contentions of each side can be combined in his new, third option. But I claim that his account cannot accommodate the central insight of one of the rival views (that of Anscombe and Davidson), and that it retains a central defect of the other (that of Goldman).[1]

I

Thalberg dubs those who take Anscombe's and Davidson's view of individuation *Unifiers*, and those who take Goldman's *Pluralists;* and this suggests that he thinks that the issue in the past has always been conceived as an issue about number of actions—one or many (p. 87). Certainly questions about numbers of actions and questions about their identities may be distinguished, since some kinds of things have instances that can be individuated but not enumerated.[2] But we can settle in a particular case whether *a* is the same action as *b*, without raising the question whether, if *a* ≠ *b*, the correct answer to *How many actions?* would be straightforwardly *Two*. So if we do separate identity questions from counting questions, identity questions comes first; and an issue about identity divides Anscombe and Davidson from Goldman before any counting questions are allowed to impinge.

We can see this using Thalberg's own example (pp. 84-5). Imagine then the Prime Minister of Acirema who presses a button, does so by moving a finger, and by doing so destroys the city of Dauphinia. Take any pair of descriptions in the series 'his moving his finger', 'his pressing the button', 'his destroying Dauphinia', construct an identity statement using those descriptions, and you will have a statement that Anscombe and Davidson hold true, Goldman false.

Thalberg's own view is that there are not 'separate, numerically distinct actions' here (p. 107). Rather the events in such a series have

¹ Thalberg, *Perception, Emotion and Action* (Blackwell, Oxford, 1977); G. E. M. Anscombe, *Intention* (Blackwell, Oxford, 1957); Donald Davidson, 'Agency' in *Agent, Action and Reason,* eds. R. Binkley et al. (University of Toronto Press, 1971) pp. 3–25; A. Goldman, 'The Individuation of Actions', *Journal of Philosophy XVIII* (1971) pp. 761–74. Otherwise unattributed page references are all to these works.
² This is a point on which absolutists and relativists about identity are agreed. See e.g. David Wiggins, *Identity and Spatiotemporal Continuity* (Blackwell, Oxford, 1967) pp. 39-40, and P. T. Geach, in many places. (To see that the point applies to (some kinds of) actions, one could use the method that Geach uses to make the point in his 'Note' on O. R. Jones's 'Identity and Countability', ANALYSIS 24.6 (1964) pp. 205-6.)

components in common. An event added on to the action of moving the finger gives the action of pressing the button, and some further events added on to that—missiles' blasting off and their razing the city, in Thalberg's example—give the action of destroying the city. What Thalberg wishes to say on the issue about identity is not always clear. He holds that the Prime Minister 'in carrying out [these things] does not perform different actions' (p. 110), which makes it sound as though he takes the actions to be the same.[1] And we appear again to be encouraged to think they are the same when he states 'I am interested in the analogues of a material object taking on . . . more parts' (p. 111). It seems as if he imagines that the very event that is his moving his finger becomes his destroying the city by the addition to it of other events—just as a tadpole may grow legs and thereby become a thing with legs. Thalberg must allow for some breakdown of analogy here, however, because he does not think that an event of missiles' blasting off is ever a part of a man's moving his finger—as the legs of a tadpole are at some time a part of it. And this reminds us that Thalberg really must deny the identity: he believes that the action of destroying the city has parts that the action of moving his finger lacks. At one point he explicitly asserts their non-identity (p. 110). And I can only suppose that when he says that the Prime Minister did not perform different actions, Thalberg has allowed his insistence on a distinction between questions of identity and questions of number to lead him to make a distinction between questions of identity and questions of difference.

Thus Thalberg is clearly on one side of one controversy. If we called the rival parties not Unifiers and Pluralists, but Identifiers and Differentiators, it would be clear which side he took. These indeed might be more accurate labels to attach to the disputants, because, despite Thalberg's suggestion that questions of number have had undue influence, the only writer who has placed any emphasis on these questions is Judith Jarvis Thomson—and she shares Thalberg's views about identities. Goldman himself says that claims about number need not be in question, and Davidson has made it clear that he thinks they are a separate issue.[2] But however that may be, Thalberg claims it as a distinctive virtue of his component approach that it enables us to refuse to say that the Prime Minister performed a large number of separate

---

[1] Talk about 'performing actions' may be like talk of 'doing things', which I discuss in the last paragraph below. If it is, then Thalberg's claim that the Prime Minister did not perform different actions is going to be false if only the Prime Minister's action was of more than one kind. This makes me think that Thalberg must mean to be saying that there weren't different particular actions, so that he is for a moment siding with the Unifiers.

[2] Judith Jarvis Thomson, 'The Time of a Killing', *Journal of Philosophy LXVIII* (1971) p. 115. (To see that she shares Thalberg's views about identities, see *Acts and Other Events*, Yale University Press, 1977, where she no longer stresses number claims.) Goldman, *loc. cit.* p. 773. Davidson, 'The Individuation of Events', p. 233 in *Essays in Honour of Carl G. Hempel*, ed. N. Rescher (Dordrecht Reidel 1969).

actions, while admitting that his actions took place at different times. So we should ask: are there good grounds for thinking that the actions were not simultaneous? (II); and will Thalberg's negative claims about number be enough to satisfy those who thought, positively, that we find only the same action described and re-described? (III).

II

Thalberg makes two related claims in support of not identifying basic actions (like someone's moving his finger) and related non-basic actions (like someone's pressing a button or his destroying a city) (pp. 102–5). The first is that the non-basic action 'takes effect' at a later time than the basic action, the second that the non-basic action cannot be said to have happened until later than the basic one.

What is it for a particular event to take effect? The question is intrinsically indeterminate. If we call an action a simply, and ask when it took effect, we have no idea how to answer—until it is said which effect of a is meant. If somebody calls an action a pressing of a button, or if he calls an action a destroying of Dauphinia, then it is clear enough which effects he has in mind: determinacy is secured in the manner of mentioning the action: he wants to know when a button was depressed, or when Dauphinia was destroyed. To say that an-action-of-kind-K took effect at t (I suggest) is to say that the effect of the action that it is revealed to have had when it is seen to be of kind K—that effect occurred at t. And of course the various effects of a single action—the effects that are revealed when we see it as of various non-basic kinds—occur at different times. Thus if we receive different answers to the questions *When did a of kind K take effect?* and *When did b of kind L take effect?,* this may be because $a \neq b$ or it may be because $K \neq L$; and the different answers that we get to the questions when his action of moving a finger and when his action of destroying a city took effect need not show that the questions are concerned with different actions.

In attempting to show that non-basic actions take longer to complete than their basic counterparts, Thalberg would have us consider a time at which the Prime Minister has moved his finger, but a time at which the city of Dauphinia still stands. At such a time, he would say, the Prime Minister has not destroyed Dauphinia. But even if one agrees that 'He has destroyed the city' would be false if spoken before the city's devastation, it is still possible to query whether that proves that

---

On the question of identities itself Thalberg actually occupies a middle ground between Goldman and Anscombe or Davidson. (Although he would agree with Goldman about the case as it has been described so far, still, if the details of what the Prime Minister did were further elaborated, then we should find pairs of action descriptions which Goldman would take to describe different actions whereas Thalberg held that they described the same action.) But since I hope to find support for the many identity claims that Thalberg and Goldman and many others would deny, the details of particular accounts need not concern me here.

his action of destroying the city has not occurred then. If (as Thalberg agrees) the Prime Minister's destroying the city is his doing something that causes its destruction, then the sentence 'He has destroyed the city' is true of him only if he has the property that anyone who has caused a certain city's destruction has. But this is a property that the Prime Minister does not come to have until something he does *has* caused the destruction. If so, he comes to have destroyed the city, and his action comes to be a destroying of the city, at a time later than the occurrence of his action that is his destroying the city. On this account of the matter, the tensed sentences are a distraction: they direct our attention away from the crucial question of when the action occurred, and towards the different question of when it can first be said to have certain properties.[1] And unless Thalberg has an account of tense that contradicts this, his case cannot rest upon our views about tensed sentences.

III

It seems reasonable to think that an action that takes a stretch of time to occur is occurring at any moment in that stretch, and reasonable too to think that if *a* is agent, then any moment at which his action is occurring is a moment when *a* is doing something. If the action that is the Prime Minister's destroying the city goes on occurring at a later time than that at which he moves his finger, then it seems he must be doing something after he moves his finger. But suppose that the Prime Minister moved his finger just once, and that he dropped dead the moment his finger left the button. That need make no difference to whether he did destroy Dauphinia. But it will make a difference to when he was doing things: he did nothing after he pressed the button in this case. So his action of destroying the city cannot have been occurring after the time that he moved his finger.

An argument very like this is used by Davidson (p. 21, but from his version we could deduce the stronger conclusion that here his destroying the city is the same as his moving his finger—see below). Thalberg says that he recognizes the force of such argument. But my version is aimed directly against Thalberg's claims, and he must reject one of my premises. He must either deny that an agent has to be doing something for the duration of any of his actions, or else allow that an agent may be doing things after he is dead.

Thalberg would respond by saying that we shall find nothing untoward in thinking that actions may continue after the agents' deaths, provided that we appreciate that actions carry on of their own momentum; they take on broader dimensions without the agent's aid;

---

[1] Cf. Jonathan Bennett's idea of *delayed properties*, 'Shooting, Killing and Dying', *Canadian Journal of Philosophy* II (1973) pp. 315–23.

parts are added to them with time.[1] But what we then have to ask Thalberg is *what* action it is that carries on after the agent is no longer active. What is it that, in his example, acquires parts when the Prime Minister's finger has left the button? The answer cannot be *His moving his finger*. Although Thalberg can say that that action occurs when the agent is doing things, he cannot say that *it* acquires a part afterwards, because he claims that actions of moving the finger finish when the finger stops moving. Again, the answer cannot be *His destroying the city*. Although Thalberg can say that that action continues after the agent stops doing anything, he cannot say that *it* acquires a part after then: he claims that the pressing of a button is the event to which a part is added to obtain his destroying of the city. But this is different from it, from his destroying of the city.

The difficulty here is a general one. It started to surface in Section I when we noticed the breakdown of analogy between the acquisition of their spatial parts by material objects and the acquisition of their temporal parts by events. Continuants take up space in the world, and they can acquire spatial parts as they persist through time. Events take up time in the world, and though their own parts may continue to happen as time progresses, we cannot make sense of events acquiring other events as parts of themselves. At least we cannot make sense of that if we acknowledge that the proper parts of events are not the same as the events of which they are proper parts. Then we cannot say of an event a that it acquires e as a part, since the event $(a+e)$ that has e as a part must be different from a (in the uncontroversial sense of different for which non-identity is sufficient). This is why actions cannot occur at a time when the agent has ceased to do anything. There can be no candidate for an event which is both someone's doing something and the same as an event containing a part that is not his doing anything. So if we were to allow that some action might go on occurring after the agent was dead, then it would have to be different from any action that occurred when he was alive and doing things. As far as I know, no-one has ever denied that there is for every action some time when it is occurring and at which the agent is doing something.

Thalberg spoke of his account's making concessions both to Unifiers and to Pluralists. But no real concession is made to the Unifiers (sc. Identifiers). If I am right, they were concerned with identity claims, and their arguments do lead to a conclusion about identities of actions that Thalberg's account rules out. In virtue of moving his finger, the Prime Minister has done everything he needs to do in order for certain other things to have been done by him. He can drop dead, and he will

---

[1] P. 107. Cf. also Monroe Beardsley, 'Actions and Events: The Problem of Individuation'. *American Philosophical Quarterly XII* (1975) pp. 263–276. The argument that follows has benefited from the Editor's criticisms.

still have done them. Thus his moving his finger must be the same as his doing these other things. It must be the same as his pressing the button, and the same as his destroying the city.

IV

I end by showing how much I think Unifiers can concede to Thalberg. There is something that Pluralists say, which Thalberg thinks it is a special feature of his account to make room for, but which, it seems to me, no Unifier ever needs to deny.

> [W]e should concede to Pluralists . . . that when he razes Dauphinia, he accomplishes more . . . . Thanks to the circumstances and surroundings in which he bends his finger, this basic action results in the devastation of Dauphinia. Such . . . consequences . . . may rank as things the Prime Minister accomplishes by crooking his finger. (pp. 106-7).

Here an agent is said to accomplish the consequences of his action; and if consequences are the things that we accomplish, then small wonder that we can accomplish several things at a single stroke: an action, like any other event, can have a large number of consequences. But even if we do not wish to speak of accomplishing consequences, the Unifier, like anyone else, can and should allow that the man who razed the city accomplished, or did, several things. The reason is that (e.g.) 'He did three things' tells us nothing about the number of his particular actions.

This is a different point about number from the one that Thalberg makes. A statement like 'He accomplished a lot', or 'He did three things', is irrelevant to the controversy about the individuation of actions not merely because (as Thalberg insists) identity claims cannot always be settled by reference to claims about number. 'He did three things' is irrelevant because it speaks about the number of the wrong sort of thing: it changes the subject—away from particular actions, to kinds of actions. We can see this in the use of parallel claims about sameness. The phrase 'did a thing' as it occurs in 'I did the same thing as you' is presumably to be construed on a par with its occurrence in 'He did three things'. But a sentence like 'I did the same thing as you' does not assert an identity between actions—at least if actions are taken to be particulars. For if there are particular actions, then it is surely a sufficient condition of difference that one action is yours, another mine. To ask how many things someone did, or to ask what he accomplished, is to enquire how many (significant) kinds of action were his, or what kinds of action were his.[1] A single action may have many effects, and

---

[1] And yet people on my side are apt to state the view by saying things like 'He only did one thing'. It is possible that a sentence like 'He did three things' can be interpreted in one of two ways: *He did three doings* or *He did doings of three kinds;* and it is possible that philosophical use (e.g. of the locution 'do *x*') has encouraged us to interpret the sentence in the former way. But on the basis of English use of 'He did something' (as well as 'I did

effects of many kinds: it may be an action of many kinds. We sometimes achieve a great deal by moving our bodies.

Corpus Christi College, Oxford © JENNIFER HORNSBY 1979

NO ARGUMENT AGAINST RAMSEY

By LAURENCE NEMIROW

EVER since Ramsey denied the distinction between the logical roles of the subjects and predicates in subject-predicate sentences ('Universals', in *The Foundations of Mathematics*, London: Routledge and Kegan Paul, 1954), some philosophers have felt challenged to redeem the distinction. Geach has claimed that the roles of subjects and predicates are distinguished by the fact that expressions for predicates, unlike subject terms, come in complementary pairs ('Assertion', *Philosophical Review* (October, 1965), p. 461; for a closely related argument, see Geach, *Reference and Generality* (Ithaca: Cornell University Press, 1962), pp. 31–34). Geach's argument has been adopted by Anscombe ('Retractation', ANALYSIS 26.2 (December, 1965), pp. 33–36) and by Strawson ('The Asymmetry of Subjects and Predicates', in *Logico-Linguistic Papers* (London: Methuen, 1971), pp. 3–40; see also, *Subject and Predicate in Logic and Grammar* (London: Methuen, 1974), pp. 5–7). And, most recently, Geach has proposed a revised version of his argument, from complementation, for the subject-predicate distinction ('Names and Identity', in *Mind and Language,* ed. S. Guttenplan (Oxford: Clarendon Press, 1975), pp. 143–146).

I will use one name, the 'No' argument, to refer to both the original and the revised version of Geach's argument. Both versions of the No argument, I will contend, are fallacious.[1]

This paper has three parts. Parts one and two evaluate the earlier and later versions of the No argument, respectively. Part three is a

---

the same thing as you'—see text), the latter is what we should predict that the sentence says. If this is right, so that Unifiers sometimes resort to dubious English to express their theses, then that may explain some of the resistance to their views.

[1] For Strawson, the asymmetry with respect to complementation between subjects and predicates closely parallels another asymmetry: Predicates, unlike subject terms, can be conjunctively and disjunctively compounded (*Subject and Predicate,* pp. 6-7). I will not discuss Strawson's argument for the latter asymmetry. But my remarks concerning the early No argument will also apply to Strawson's argument with only minor, systematic changes.

description of the fallacy which underlies both versions of the No argument.

                                          I

Ramsey acknowledged the 'grammatical' distinction between subjects and predicates ('Universals', p. 117). But he denied the distinction between the logical roles of subjects and predicates (for short, 'the logical subject-predicate distinction') on the ground that those roles are perfectly symmetrical. To substantiate his claim of symmetry, he argued that any subject-predicate sentence can be rewritten with the roles of subject and predicate reversed, and without a change of meaning. For example, 'Socrates is wise' can be rewritten, 'Wisdom is a characteristic of Socrates'. Since the same proposition is expressed with either 'Socrates' or 'wisdom' as subject, there is no distinguished logical subject of the original sentence 'Socrates is wise'. In general, a predicate can be construed as a name for a quality, just as we understand a subject as naming an individual. So it is wrong to assume that in a sentence containing a subject and a predicate, 'these two terms must be functioning in different ways' ('Universals', p. 116; for a further development of this theme, see 'Universals', pp. 120–134).

Geach's No argument champions the logical subject-predicate distinction. Before stating the first version of that argument, I will take note of some terminological distinctions. In *Reference and Generality*, Geach distinguishes the meaning of 'name' from the meaning of 'subject'. A subject is a name 'in use' in a subject-predicate sentence; or, conversely, a name is a 'possible subject'. He also assigns different meanings to 'predicable' and 'predicate'. A predicate is a predicable in use in a subject-predicate sentence; a predicable is a possible predicate. Hereafter, I will follow Geach's use of these terms. (Thus, when I speak of the logical role of a subject, I mean the logical role of a name in use in a subject-predicate sentence).

The early version of the No argument (in 'Assertion') has two premises, which are expressed with more precision in *Reference and Generality*.[1] The first is that predicables come in complementary pairs, so that, as Geach writes, 'by attaching such a pair to a common subject we get a contradictory pair of propositions' (*Reference and Generality*, p. 32). The second premise is that names never come in contradictory pairs; that is, 'we never have a pair of names so related that by attaching the same predicate to both we always get a pair of contradictory propositions' (ibid.).

---

[1] In *Reference and Generality*, Geach uses the two premises to argue for a different conclusion from the conclusion of the argument, in 'Assertion', that there is a distinction between the logical roles of subjects and predicates. The conclusion in *Reference and Generality* is the distinction of names and predicables: One unambiguous expression cannot be a predicate in one sentence and a subject in another. I am directly concerned only with the logical subject-predicate distinction.

The first premise is apparently true. In *Reference and Generality* (pp. 32-33) Geach gives what he calls a 'formal' argument for the second premise. That argument has a premise of its own: Predicables can be conjunctively compounded; that is, given any predicables P and Q, there is a predicable $(P \& Q)$, such that if c is a name, $\ulcorner(P \& Q)(c)\urcorner$ and $\ulcorner P(c) \& Q(c)\urcorner$ are logically equivalent.

The new premise yields the conclusion that names do not come in complementary pairs. For suppose, to the contrary, that c and c' are complementary names. Let $\ulcorner P(c)\urcorner$ and $\ulcorner Q(c)\urcorner$ be subject-predicate sentences which have opposite truth-values, but c as their common subject. Now either $\ulcorner P(c')\urcorner$ or $\ulcorner Q(c')\urcorner$ must be false, since either $\ulcorner P(c)\urcorner$ or $\ulcorner Q(c)\urcorner$ is true. Hence $\ulcorner Q(c') \& P(c')\urcorner$ is false. But $\ulcorner Q(c') \& P(c')\urcorner$ is equivalent to $\ulcorner(Q \& P)(c')\urcorner$ which is true because $\ulcorner(Q \& P)(c)\urcorner$ is false. This is absurd; so there are no complementary pairs of names.

The early No argument is now almost complete. From the premise, proved above, that names do not come in complementary pairs, together with the observation that predicables do, Geach concludes to the logical subject-predicate distinction ('Assertion', p. 461).

Crucial to evaluating this version of the No argument is a clear understanding of the second premise. The second premise does not say that names cannot have complements, but that the complement of a name cannot itself be a name. Geach's formal argument demonstrates only the latter, weaker conclusion because it assumes the equivalence of $\ulcorner(P \& Q)(c')\urcorner$ and $\ulcorner P(c') \& Q(c')\urcorner$, which we cannot know unless we know that c' is a name. For we characterised the result of conjunctively compounding predicables only by saying what happens when a compound predicable is attached to a *name*.

In fact, it is not difficult to define a language which satisfies the premises of Geach's formal argument, and in which names have complements. Suppose that L is a formal language. By the phrase 'a predicable of L', I shall mean any formula which yields a closed sentence when attached to a single occurrence of an individual constant. Assume that the predicables of L are closed under complementation and conjunctive compounding (i.e., for any predicables P and Q, there are predicables $(P \& Q)$ and $-P$ such that $\ulcorner(P \& Q)(c)\urcorner$ is true in a structure just in case both $\ulcorner P(c)\urcorner$ and $\ulcorner Q(c)\urcorner$ are true, and $\ulcorner(-P)(c)\urcorner$ is true just in case $\ulcorner P(c)\urcorner$ is false). I shall now explain how to extend L to a language L', so that the predicables of L' are still closed under complementation and conjunctive compounding, and so that the names of L' have complements.

The primitive vocabulary of L' consists of the primitive vocabulary of L, together with a new and distinct symbol c' for each individual constant c of L. c' will be the complement of c. A formula of L is any sequence of symbols which results from a formula of L by replacing

zero or more occurrences of individual constants by their complements. Each formula of L' thus results from a unique formula of L, which I shall call its 'source'.

The satisfaction relation for L' is determined by the satisfaction relation for L. A sequence satisfies a formula of L' in a structure if and only if either (a) the formula has an even number of occurrences of complements of names, and the sequence satisfies the formula's source, or (b) the formula has an odd number of occurrences of complements, and the sequence fails to satisfy the formula's source.

By 'a predicable of L'', I mean any formula of L' whose source is a predicable of L. Relative to L', and relative to this notion of predicable, the premises of Geach's early No argument hold. For the predicables of L' are closed under complementation and conjunctive compounding, as a formal exercise will show. And in L', names have complements: $\ulcorner P(c)\urcorner$ is true in a structure if and only if $\ulcorner P(c')\urcorner$ is false, for each predicable P and name c. Thus, Geach's argument does not demonstrate that names cannot have complements. The argument only shows that the complement of a name is not a name.[1]

So a premise of the early No argument is this: There are no complementary pairs of names. It is not: A name cannot be complemented. With this in mind, I can now describe the fallacy in the argument. It is the fallacy of trying to distinguish *objects* of sort A from *objects* of sort B by demonstrating a property of the *class* of A's which is not a property of the *class* of B's. Granted that Geach's formal argument shows that the class of names is not closed under complementation of any of its elements, and granted that the class of predicables is closed under complementation, it does not follow that names and predicables in subject-predicate sentences have distinct logical roles. All that really follows is that no predicable and its complement are both names.

Consider an analogy. Fix an object O. Call a predicable an 'O predicable' if it is truly predicable of the object O. It is obvious that the complement of an O predicable is not an O predicable. Now if it were legitimate to argue that names and predicables in subject-predicate sentences have distinct logical roles, simply on the ground that

---

[1] It may be objected that names can have complements in L' only because L' is an impoverished language. For it will contain neither conjunction nor disjunction. E.g., there will be no sentential connective '&' such that for all sentences S and T, $\ulcorner S$ & $T\urcorner$ is true in a structure just in case S and T are both true. Now it is true that L' will not have symbols for conjunction or disjunction. But it is not necessarily impoverished. Obviously, L' is no more impoverished than L, since L' contains L. (Thus, L' will contain symbols for conjunction and disjunction with respect to sentences of the original language L, assuming that L contains conjunction and disjunction). Furthermore, conjunction and disjunction can be expressed in L', notwithstanding the fact that there are no symbols for conjunction and disjunction. For there are obvious, mechanical procedures for expressing the conjunction and disjunction of any two L' sentences. And conjunction and disjunction symbols can be introduced as abbreviations for the result of applying those procedures.

predicables, unlike names, come in complementary pairs, then it would be equally legitimate to argue that O predicables and predicables have distinct logical roles in sentences. But that is absurd because, of course, an O predicable is a predicable.

Here is Geach's statement of the early No argument:

> What distinguishes predicates from subjects, I suggest, is . . . that by negating a predicate we get the negation of the proposition in which it was originally predicated. (Plainly, there is nothing analogous for subject terms). ['Assertion', p. 461].

If this argument were correct, the following argument would be correct:

> What distinguishes shoes from right shoes is that shoes come in left-right pairs. (Plainly, there is nothing analogous for right shoes).[1]

It may seem possible to save the No argument by finding some appropriate feature of predicables which is not a feature of names (rather than a feature of the class of predicables which is not a feature of the class of names). A likely candidate is the feature of having a complement *and* having the complement's role: subjects would be distinguished from predicates by the fact that a predicable, unlike a name, both has a complement and retains the complement's role when replacing the complement in a sentence. (Strawson strongly hints at such an argument in 'The Asymmetry of Subjects and Predicates,' p. 98).

This modified version of the early No argument is inconclusive. For it presupposes an understanding of the roles of a sentence's subject and predicate. But those roles are the very subject of dispute.

A proponent of such a modified No argument must provide a method for assessing the roles of a name and its complement, in order to justify the assumption that a name cannot retain the role of its complement when replacing the complement in a sentence. Geach has pointed out to me that, on page 59 of *Reference and Generality*, he does state a method for distinguishing genuine subjects from terms which merely resemble subjects grammatically. Subjects are distinguished by the fact that they behave correctly with respect to conjunctively and disjunctively com-pounded predicables; thus, $\ulcorner(P \ \& \ Q)(c)\urcorner$ must be equivalent to $\ulcorner(P(c)$ $\& \ Q(c))\urcorner$, for logical subject c. This precludes a name's complement from having the role of subject because, as Geach's formal argument shows, complements of names do not behave correctly with respect to conjunctively compounded predicables.

---

[1] Geach's argument in *Reference and Generality* for the distinction of names and predicables (see note 2) is similarly invalid. He argues, 'No names come in contradictory pairs; but all predicables come in contradictory pairs; therefore no name is a predicable' (p. 33). He might as well argue, 'All shoes come in left-right pairs; no right shoes come in left-right pairs; therefore no right shoe is a shoe.'

Now this line of reasoning shares an important and questionable presupposition with Geach's formal argument. They both presuppose that the complement of a name must invert the truth value of a sentence formed from a compound predicable: $\ulcorner(P \ \& \ Q)(c')\urcorner$ is equivalent to $\ulcorner-(P \ \& \ Q)(c)\urcorner$, if c is a name and c' is its complement.

Suppose that, in defiance of this presupposition, we were to redefine the semantics for complements of names as follows: Formulas containing only non-compound predicables are treated as in L' above. In other words, the semantical value of such a formula is determined by whether it has an even or odd number of occurrences of complements. But the semantical values of sentences containing compound predicables are determined in the usual way from the values of simpler formulas. For example, a sentence of the form $\ulcorner(Q \ \& \ P)(c')\urcorner$ is true just in case $\ulcorner Q(c')\urcorner$ and $\ulcorner P(c')\urcorner$ are both true. Under this modified semantics for complementation, it would be false to assume, as Geach's formal argument does, that complements of names invert the truth values of sentences formed from compound predicables. Moreover, both names and their complements would behave correctly with respect to compound predicables; hence, they would satisfy Geach's criterion for a term's having the role of a subject.

Geach might contend that we should not count a term as the complement of a name unless it inverts the truth value of a sentence formed from any predicable, not just from a non-compound predicable, just as the complement of a predicable must invert the truth value of a sentence formed from any name. I can reply, however, that the complement of a name should be required only to invert the truth value of a sentence formed from a non-compound predicable, just as—as Geach would agree—the complement of a predicable is only required to invert the truth value of sentences with non-compound terms in subject position.

Regardless of how this issue is to be resolved, it reveals the importance of a fact which has been insufficiently emphasised: complementation is defined relative to some class which does not remain constant. Complementation for predicables is defined relative to the class of names, whereas complementation for names is to be defined relative to some class of predicables. Consequently, an asymmetry between names and predicables with respect to complementation would not signify a distinction between names and predicables with respect to one constant property. This fact is especially significant because Geach's No argument crucially presupposes that complementation for names is defined with respect to the class of all predicables, as opposed to the class of non-compound predicables. And this reinforces my previous conclusion that the asymmetry of names and predicables with respect to complementation can signify only an asymmetry between the class

of names and the class of predicables (but not between the class of names and certain sub-classes of predicables, such as the class of O predicables, or the class of non-compound predicables). I will return to this topic in part three.

II

Geach's recent version of the No argument (in 'Names and Identity', pp. 143–147) purports to demonstrate that names do not have complements of any kind, and not merely that names do not come in complementary pairs. Geach relies on a different concept of predicable from the one I adopted in constructing a language with complements of names, in section one. My construction yielded a language with complements only on the assumption that a predicable results from detaching a single occurrence of a name from a sentence. In contrast, Geach now assumes that a predicable could result from detaching several occurrences of the same name from a sentence. His assumption implies that the predicates of the sentence 'Sam loves Sam' would not be only 'Sam loves ——' and '—— loves Sam' but also '—— loves ——', whereas I countenanced only the first two forms.

The thesis that names do not have complements now follows directly from Geach's assumption about predicables. If 'Sam″ were the complement of 'Sam', then 'Sam loves Sam' and 'Sam′ loves Sam' would have opposite truth values, since these sentences result from attaching the same predicable, '—— loves Sam', to 'Sam' and 'Sam″ respectively. Also 'Sam′ loves Sam' and 'Sam′ loves Sam″ would have opposite truth values, as they result from attaching the same predicable, 'Sam′ loves ——', to 'Sam' and 'Sam″ respectively. Thus, 'Sam loves Sam' and 'Sam′ loves Sam″ would share a truth value, since they both differ in truth value from 'Sam′ loves Sam'. But this is absurd because they result from attaching the same predicable, '—— loves ——' to 'Sam' and 'Sam″ respectively. Hence, names cannot have complements. Since predicables do have complements, Geach concludes to the logical subject-predicate distinction: 'I conclude [that] in a two-term predicative proposition, only one of the two terms has a predicative role' ('Names and Identity', p. 146).

This version of the No argument is not compelling. For it fails to prove its point for languages in which predications are unambiguous, that is, where there is no opportunity to analyse one formula into subject and predicate in more than one way. The argument fails, for example, in lambda notation.

In lambda notation, a subject-predicate sentence, ordinarily represented by $\ulcorner M(a) \urcorner$, is represented by $\ulcorner \lambda x(M(x))(a) \urcorner$, where $\ulcorner \lambda x(M(x)) \urcorner$ is the predicate, and a is the subject. The sentence 'Sam loves Sam' can be represented by any of these formulas:

'λx(x loves x)(Sam)' or 'λx(λy(x loves y)(Sam))(Sam)' or 'λy(λx(x loves y)(Sam))(Sam)'

each of which contains only one of the following predicates:

'λx(x loves x)', 'λx(λy(x loves y))(Sam)' and 'λy(λx(x loves y)(Sam)';
i.e. ' —— loves ——', '—— loves Sam', and 'Sam loves ——'.

The multiple representability, within lambda notation, of 'Sam loves Sam' confirms Geach's claim that 'Sam loves Sam' admits the three different subject-predicate readings. But we can use lambda notation to sidestep Geach's argument, based on that multiple analysis, against the possibility of complementing names. For predications formulated within lambda notation do not admit of various subject-predicate readings: a subject-predicate sentence ⌜λx(M(x))(a)⌝ has a unique decomposition into predicate ⌜λx(M(x))⌝ and subject a. Moreover, lambda notation conforms to my original assumption that a subject-predicate sentence consists of a predicable attached to a single occurrence of a name. So a language in lambda notation can be enriched with complements of names, following either of the two procedures described in section one. And adopting a lambda notation as the base language imposes no cost in expressibility.

My strategy for introducing complements of names, then, is to disambiguate predications. I concede that some predications in English admit of different subject-predicate readings, and that some formal languages may contain this feature too. But I insist that the feature can be eliminated, without loss of expressibility. Of course, the feature is not important for ordinary logic, since the various readings of any one predication are obviously logically equivalent. But it becomes important once we insist on enriching a language with complements of names. The language must be 'disambiguated' first.

III

Neither version of Geach's No argument shows that names cannot have complements. The early version does not even claim to do so; and the recent version does not succeed in demonstrating as much. But even if it were established that names cannot have complements, the logical subject-predicate distinction would not follow directly. For complementation is always defined relative to a certain class. And an asymmetry between two classes with respect to complementation defined relative to one another does not imply an asymmetry between the elements of those classes, as an example from set theory will show.

I will say that sets x and y are complements relative to a class C if C is disjoint from the intersection of x and y, and if C is a subclass of the union of x and y. Now every set x has a complement relative to the set

of natural numbers N, namely, N − x. In contrast, no natural number has a complement relative to the class of all sets, since that class is the union of no two sets. It does not follow, however, that the elements of N are not sets, or that terms for those elements are not terms for sets. To the contrary, elements of N *are* sets, and terms for its elements *are* terms for sets.

This example does not invoke a new sense of 'complementation'. For in the example, as in our discussion of names and predicables, the relation of complementation holds between objects in their capacities as principles of collection: two things are said to be complementary with respect to some class if there is no member of that class which they both collect and if, jointly, they collect all the members of that class. Thus, two sets are complementary with respect to the class N if there is no natural number they both collect and if, together, they collect all natural numbers. And two predicables are complementary with respect to the class of names if there is no name they both collect and if, together, they collect all names. In all cases, complements are allowed to collect common objects which are outside the class with respect to which complementation is to be defined.

The example shows that, given this notion of complementation, an argument of the following sort is not valid:

All A's have complements relative to the class of B's.

No B's have complements relative to the class of A's.

A's have a property which B's lack.

The two premises in this argument actually entail, not a distinction between A's and B's, but only an asymmetry between the class of A's and the class of B's.

At best, the No argument demonstrates an asymmetry between the class of names and the class of predicables with respect to complementation defined relative to one another. Although the existence of this asymmetry may be of fundamental importance, it is no cause for concern to the followers of Ramsey.[1]

---

[1] I am grateful for the valuable advice of Nancy Cartwright, Peter Geach, and Joanne Kadish.

Stanford University   © Laurence Nemirow 1979

CAUSATION, COMPATIBILISM AND NEWCOMB'S PROBLEM

By Don Locke

IN 'How Not to Make a Newcomb Choice' (ANALYSIS 39.1, January
1979) Andre Gallois makes a number of criticisms of my attempted
solution of the Newcomb problem (ANALYSIS 38.1, January 1978).
These reduce, however, to two general metaphysical issues which
happily can be discussed without rehearsing the details either of the
problem or of my argument. Nevertheless one, the role of reverse
causation, is an important factor in concealing what seems to me the
correct solution.

I

I argued that the possibility of reverse causation is irrelevant to
the problem since, even if my present choice in some sense causes the
Predictor's past behaviour, at least it cannot alter it. Gallois replies
that this is of a piece with the standard fatalist argument that you cannot
alter the future. But we can distinguish between altering the present,
past or future in the sense of making it other than it actually is, was or
will be, which is of course impossible; and altering it in the sense of
making it other than it would otherwise be, i.e. without the intervention
of the altering event. In the latter sense you can alter the future and
even the present, if it is long enough and you are quick enough, but
not the past—whatever I do now, whether I do it or not, the past will
remain as it is—and it was this that I had in mind when I dismissed as
incoherent the suggestion that one might 'actually alter the past, make
it other than it was' (my p. 18). What this means, I think, is that if
some present event B is genuinely to cause some past event C, B must
itself be caused by some event A simultaneous with or prior to C (the
matter is well discussed in J. L. Mackie, *The Cement of the Universe*,
Chap. 7). But the important point is that nothing I do now can make a
difference to the Predictor's past behaviour, whereas if that behaviour
lies in the future, what I do now may make all the difference in the
world.

There is, therefore, a certain irony in Gallois's charge that my argu-
ment 'is simply a standard argument for fatalism applied to the past
instead of the future' (p. 50), since the past is precisely where fatalism
can be expected to apply. The problem of fatalism, rather, is caused by
extending a commonplace about the past ('either there was a sea-battle
yesterday or there was not: there is nothing we can do about it now') to
the future ('either there will be a sea-battle tomorrow or there will
not: there is nothing we can do about it now'). I think it is obvious,
without further argument that this same misunderstanding, about the

relative alterability of past and future, also underlies Gallois's objections to Schlesinger's well-wisher argument.

Incidentally Gallois does not explain, any more than I did, how he understands reverse, or ordinary, causation, but there is the implication that A will cause B provided A is a sufficient condition of B. If that is all there is to it then reverse causation becomes an exceedingly common phenomenon, since A is a sufficient condition of B wherever B is a necessary condition of A: my death, for example, will be a cause of my birth. If that is your concept of reverse causation, you are welcome to it.

II

Gallois is right to point out that, in terms of my argument, the sense in which it is impossible for me to choose differently once the Predictor has predicted differs from the sense in which it is impossible ever to act contrary to a genuinely infallible predictor. I therefore do not embrace the familiar modal fallacy, but rather the argument that once 'the Predictor's past prediction is fixed, the Chooser's choice is fixed as well' (Gallois p. 52; there is, incidentally, a serious misrepresentation in the interpolation Gallois makes to the passage quoted at the top of his p. 53, as a glance at the original context, my p. 21, will reveal). Gallois argues that it is nevertheless causally possible for me to choose differently, since it is causally possible for the Predictor to have made a different prediction. But this again ignores the unalterability of the past: it may have been causally possibly *then* for the Predictor to have acted differently, but it is not causally possible *now,* so neither is it causally possible for me to act differently now.

Gallois then argues that even if a different choice is in some sense causally impossible, this does not prevent the choice from being free, without some further argument against compatibilism. I have argued against compatibilism elsewhere (e.g. 'Three Concepts of Free Action', *Proceedings of the Aristotelian Society,* Supplementary Volume XLIX, 1975) but there is no need to repeat such arguments here, since I am content with the conclusion that it is causally impossible for me to choose in any other way. Gallois, however, is as cavalier about compatibilism as he is about causation. No compatibilist, to my knowledge, has wanted to claim that *every* event which is causally determined can also be classified as free. Instead it is usually conceded that a choice, for example, is free only if it is caused by the agent's own desires, or some such. So to show that my choice is a free one, even in a compatibilist sense, it would be necessary to show that it is not in fact caused either by the Predictor's action, or by some independent third thing which causes them both—which would seem to me the two most obvious explanations, were a Newcomb predictor ever actually to appear.

University of Warwick © DON LOCKE 1979

REID AND MILL ON HUME'S MAXIM OF CONCEIVABILITY

By ALBERT CASULLO

THE concepts of conceivability and inconceivability have played an important role in the history of British philosophy. The reason for their importance is made explicit by Hume in the following passage:

> 'Tis an establish'd maxim in metaphysics, *That whatever the mind clearly conceives includes the idea of possible existence,* or in other words, *that nothing we imagine is absolutely impossible.* We can form the idea of a golden mountain, and from thence conclude that such a mountain may actually exist. We can form no idea of a mountain without a valley, and therefore regard it as impossible. (*A Treatise of Human Nature,* ed. Selby-Bigge, p. 32).

This maxim consists of two principles which are logically independent of each other: (1) whatever is conceivable is possible; and (2) whatever is inconceivable is impossible. (I shall refer to the first as the principle of conceivability and to the second as the principle of inconceivability). One can consistently hold either of these principles while rejecting the other.

Thomas Reid thought that his predecessors were mistaken in attaching great significance to the concepts of conceivability and inconceivability. And in Essay 4, Chapter 3 of his *Essays on the Intellectual Powers of Man,* he offers several arguments against the principle of conceivability. Although they are not explicitly directed at the principle of inconceivability, some of them are of a general enough nature to call both principles into question. John Stuart Mill shared Reid's assessment of the philosophical significance of the concepts of conceivability and inconceivability. And in Book 2, Chapters 5 and 7 of *A System of Logic,* he offers arguments against the principle of inconceivability, although not against the principle of conceivability. The primary concern of this paper will be to examine whether Reid and Mill were successful in calling Hume's maxim into question.

I

Reid's first objection to Hume's maxim rests on considerations about what is to conceive a proposition. He begins by claiming that 'Whatever is said to be possible or impossible is expressed by a proposition', and goes on to argue that to conceive a proposition 'is no more than to understand distinctly its meaning' (Reid, *Essays on the Intellectual Powers of Man,* ed. Brody, M.I.T. 1969, p. 431). Given this interpretation of the phrase 'to conceive a proposition,' Reid is quickly able to dispose of the principle of conceivability simply by pointing out that one can

understand a demonstrably false mathematical proposition as distinctly
as one understands any true mathematical proposition. Reid goes on to
consider several other possible interpretations of this phrase but argues
that none of them yields a defensible version of the principle of con-
ceivability. His second argument is to point out that 'Every proposition,
that is necessarily true, stands opposed to a contradictory proposition
that is impossible; and he that conceives one, conceives both' (ib. p. 433).
His third argument draws attention to the fact that 'Mathematicians
have, in many cases, proved some things to be possible and others to be
impossible, which, without demonstration, would not have been
believed; yet I have never found that any mathematician has attempted
to prove a thing to be possible, because it can be conceived; or
impossible, because it cannot be conceived', (ib. p. 433-4). For example,
he claims that although it is conceivable that some number may have
the same ratio to another as the side of a square has to its diagonal, this
has been demonstrated to be impossible. Reid's final argument is that
in 'reductio' demonstrations one is asked to conceive of a proposition
which one goes on to prove is not possible (ib. p. 434).

The plausibility of Reid's first two arguments depends on an am-
biguity in the phrase 'to conceive (or understand) a proposition'. This
phrase can mean either (1) to understand the sentence which expresses
the proposition in question; or (2) to conceive of the state of affairs
described by the proposition. Reid, however, never clearly distinguishes
these two different meanings. He begins his discussion of the principle
of conceivability by stating that what is possible or impossible is
described by a proposition. If we agree with Reid that what is possible
or impossible is described by a proposition, then the principle would
read: if that which is described by a proposition (presumably, a state of
affairs) is conceivable then it is possible. But Reid does not go on to
discuss what it is to conceive of the state of affairs described by a prop-
osition. Instead, he shifts his attention to the other sense of 'to conceive
a proposition' and argues that to conceive a proposition is simply to
understand distinctly its meaning. This will not do, however, since the
most plausible reading of the principle of conceivability does not state
that one cannot understand a sentence or phrase which describes an
impossible state of affairs. This is clear, for example, from Hume's
discussion of the golden mountain. He concludes that a golden mountain
may exist on the grounds that he can form an idea of such a state of
affairs and not because he can understand the meaning of the phrase
'golden mountain'. Contemporary epistemologists have also stressed
that understanding the meaning of a sentence is not the sense of con-
ceivability relevant to epistemological issues.[1] Therefore, Reid's worries

[1] See footnote [1] overleaf.

over what it is to conceive a proposition and his subsequent criticisms miss the point of Hume's maxim.

It should be emphasised that a defender of Hume's maxim need not deny that one can conceive of an impossible state of affairs in the sense of understanding a sentence or phrase describing that state of affairs. For example, a defender of Hume can admit that he understands the phrase 'round square' without admitting that round squares are possible. For the fact that one can understand the phrase does not entail that one can also conceive (i.e., form an image) of the state of affairs it purports to describe. In fact, a person subscribing to Hume's maxim must hold that he can understand sentences which describe impossible states of affairs. For it is only by understanding such a sentence that one knows which state of affairs one must try to conceive and can discover that it is inconceivable. The relationship between understanding a sentence and conceiving of the state of affairs which it describes is very much like the relationship between understanding a recipe and actually carrying out the directions. One can understand what one has to do without trying to do it or being able to actually do it. Understanding the recipe only tells one what to do. Similarly, understanding a sentence tells one *what* state of affairs one must conceive. But whether or not one can actually conceive it is another question.[2]

The considerations I have employed to meet Reid's first argument can also be used to meet the second. It is very plausible to argue that if one understands the sentence 'Some circle is round', then one also understands the sentence 'Some circle is not round'. But the situation is not the same in the case of what a person can conceive. It is quite evident that there are many states of affairs which a person can conceive but whose negation he cannot. One can, for example, conceive a round circle but not a circle which is not round. It might be argued that the philosophers whom Reid is criticizing held that to understand a sentence is to conceive of the state of affairs which it describes. But if so, this would only show that their account of what it is to understand a sentence is mistaken. For as was pointed out above, one must be able to understand sentences which describe inconceivable state of affairs at least to the extent that one knows what he has to try to conceive.

[1] For example, Panayot Butchvarov explicitly states in *The Concept of Knowledge* (Evanston, Ill.: Northwestern University Press, 1970), p. 81, that 'Inability to understand, in the sense in which philosophers appeal to it when determining the absolute impossibility of something, is the *unthinkability* of what, in a sense, is *understood* to be described by the proposition, not the literal nonunderstanding of that proposition.' In his discussion of *a priori* knowledge, Roderick Chisholm maintains that 'To "understand" a proposition, in the sense intended, then, it is not enough merely to be able to say what *sentence* in your language happens to express that proposition. The proposition must be one that you have contemplated and reflected upon.' See *Theory of Knowledge,* 2nd ed. (Englewood Cliffs: Prentice-Hall, Inc. 1977), p. 41.

[2] Richard Taylor has emphasized the importance of this distinction for understanding the Ontological Argument. See his Introduction to *The Ontological Argument,* edited by A. Plantinga (Garden City, N.Y.: Doubleday & Company, Inc. 1965), pp. xiii–xiv.

Although Hume's principles, as stated by him, are open to Reid's third argument, I want to suggest that by invoking his distinction between intuition and demonstration, they can be revised so that they are no longer open to these objections (see Hume, p. 70). The distinguishing feature of propositions knowable by intuition is that the mind can 'see' the relationship between the ideas involved. In the case of propositions knowable only by demonstration, the mind does not have this ability and, consequently, demonstration is necessary to establish this relationship. It seems both reasonable and consistent with Hume's thinking to suppose that if demonstration is necessary to determine the truth or falsehood of certain propositions, it would also be necessary to determine their possibility or impossibility.

Employing Hume's distinction, his principles can be revised in the following manner. A proposition which is knowable by intuition is necessarily true if and only if the state of affairs described by its denial is inconceivable. A proposition which is knowable only by demonstration is necessarily true if and only if it is derivable from other propositions whose denials describe inconceivable states of affairs, using only principles of inference whose denials are inconceivable. Since a necessary proposition is one whose denial describes an impossible state of affairs, it follows that a proposition describes an impossible state of affairs if and only if *either* the state of affairs is inconceivable *or* the proposition's denial is derivable from propositions whose denials describe inconceivable states of affairs. The most important consequence of this revision is that the conceivability of a state of affairs does *not* in general entail its possibility. It entails this only if the proposition describing the state of affairs is knowable by intuition.

This two-fold account of the relationship between what is conceivable and inconceivable and what is possible and impossible will allow us to meet Reid's third objection.[1] The reason why mathematicians do not usually appeal to what they find conceivable or inconceivable in trying to establish that certain states of affairs are possible or impossible is that they are usually concerned with theorems rather than with the elementary truths of their area of investigation. Since such theorems are knowable only by demonstration, one would establish their necessity by appealing to the fact that they are derivable from other propositions whose denials describe inconceivable states of affairs and one would not attempt to show that the states of affairs described by their own denials are inconceivable. Similarly, when a mathematician proves something to be impossible which is conceivable, the proposition in question is always known only by demonstration and not by intuition. Consequently, the appeal to what one finds conceivable or inconceivable

---

[1] It will also allow us to meet a contemporary variant of Reid's objection. See my 'Conceivability and Possibility', *Ratio* 17 (1975): 118–121.

is not applicable in such cases. One establishes the impossibility of such a state of affairs by demonstration.

Reid's fourth objection, that one must conceive of propositions which one proves to be impossible in 'reductio' demonstrations, does not conflict with our revised Humean principles. For as we saw, not all conceivable states of affairs are possible. Some can be shown to be impossible by demonstration. And it is precisely in those cases where one has to employ the 'reductio' method to demonstrate the impossibility of a state of affairs that the state of affairs would be conceivable. If it were not conceivable, no demonstration of its impossibility would be needed.

II

The second set of objections to Hume's maxim which I shall examine are those offered by John Stuart Mill. Mill is primarily concerned with the principle of inconceivability and his arguments are directed mainly against his contemporaries William Whewell and Herbert Spencer rather than against Hume. Although the claims of Whewell and Spencer about the status of the principle of inconceivability are different from Hume's, the objections which Mill raises against them are general enough to call Hume's version into question. Our primary concern will be to see whether Mill has succeeded in calling Hume's maxim into question rather than to examine the theories of Whewell and Spencer.

Mill begins by pointing out that two senses of 'inconceivability' must be distinguished:

> By inconceivableness is sometimes meant, inability to form or get rid of an *idea;* sometimes, inability to form or get rid of a *belief* . . . . To illustrate the difference, we will take two contrasted examples. The early physical speculators considered antipodes incredible, because inconceivable. But antipodes were not inconceivable in the primitive sense of the word. An idea of them could be formed without difficulty: they could be completely pictured to the mental eye. What was difficult, and as it then seemed, impossible, was to apprehend them as believable. The idea could be put together, of men sticking on by their feet to the underside of the earth; but the belief *would* follow, that they must fall off. Antipodes were not unimaginable, but they were unbelievable. (J. S. Mill, *Collected Works,* ed. Robson, vol. VII, p. 269).

He goes on to argue that unbelievability cannot be a criterion of impossibility since what one finds unbelievable varies from person to person. Although this point is byond question, it does not affect Hume's maxim. For it is clear from his examples that Hume is employing 'inconceivability' in the 'primitive sense' of the word.

Mill goes on to examine the principle that whatever is inconceivable in the primitive sense is impossible. His primary argument against this principle consists in offering several examples of 'propositions now

known to be false or groundless, but whose negative was once found inconceivable: such as, that in sunrise and sunset it is the sun which moves; that gravitation may exist without an intervening medium; and even the case of the antipodes' (ib. p. 274).

In response to this argument, I want to suggest that two of Mill's counter-examples confuse the senses of 'inconceivable' he distinguishes at the beginning of his discussion. The third will be shown to be only an apparent counter-example.

First, consider the case of the antipodes. Mill explicitly states (in the above passage) that even the earlier philosophers of nature had no trouble forming the idea of men on the underside of the earth. The antipodes were conceivable in the primitive sense even to them. However, although they could form such an idea, they could not bring themselves to believe that such a state of affairs could actually occur. For they were unable to explain how these men would avoid falling from the earth. This phenomenon could not be accounted for in their framework of beliefs about the nature of the universe. But not being able to account for a phenomenon is different from not being able to conceive of it. The former has to do with the systematic explanation of a state of affairs rather than with forming an idea of it. Therefore, the correct account of this situation is that changes in our explanatory framework affect our beliefs about what is true of our world. There is no evidence that such changes affect what we can conceive in the primitive sense of the word. And Mill's claim that the case of the antipodes provides such evidence is incompatible with his own description of the situation.

Mill's second counter-example can be dealt with in a similar fashion. As people uncritically view the world around them, the belief that it is the sun which moves rather than the earth comes to them very naturally. It is often very difficult to convince a child that it is the earth which moves and not the sun. And it was even more difficult to convince people who lived at the time of Copernicus since their belief was supported by the accepted authorities of the time. Nevertheless, it is quite implausible to maintain that although such people could imagine the sun moving around the earth they could not reverse the situation and imagine the earth moving around the sun. Surely, no one needed to study astronomy and to compare the relative merits of the Copernican as opposed to the pre-Copernican theories of the solar system in order to imagine such a state of affairs. What is probably true is that although people before Copernicus could easily imagine such a state of affairs, they could not bring themselves to believe that it was the case. They could not believe that the earth moved around the sun because this belief would have been incompatible with their other beliefs about the nature of the universe. And it was only after those other beliefs were changed that they could accept this one. But it is not so obvious that

changes in our beliefs about the nature of the universe have any bearing on what one is able to conceive or imagine.

Mill's third counter-example is that although the Cartesians were unable to conceive how two bodies at a distance could affect each other without some intervening medium or underlying mechanism, we now find this perfectly conceivable. But, surely, in this case the appearance of a change in what is conceivable is the result of the fact that the concept of causality employed by the Cartesians differed from that employed by Mill. Thus, the state of affairs to be conceived under the description of one body causing another to move is different in each case. The Cartesians could conceive of two bodies in isolation moving toward each other as well as any post-Cartesian could. But what a Cartesian would deny is that one of these bodies had caused the other to move. Hence, Spencer argued as follows against Mill and in defence of the Cartesians:

> If, however, an astronomer avowed that he could conceive gravitative force as exercised through space absolutely void, my private opinion would be that he mistook the nature of conception. Conception implies representation. Here the elements of the representation are the two bodies and an agency by which either affects the other. To conceive this agency is to represent it in some terms derived from our experiences—*that is, from our sensations.* (Herbert Spencer, *The Principles of Psychology*, vol. II, p. 409).

Since the Cartesian analysis of the state of affairs '*A* moves *B*', where *A* and *B* are two entities separated in space, involves three entities—*A*, *B*, and an 'agency'—no Cartesian would describe a state of affairs consisting solely of one body moving toward another as a case of one body causing another to move. This, however, does not entail that a Cartesian could not form an image of two bodies is isolation moving toward one another.

Mill, on the other hand, can argue that he is able to conceive of one body causing another to move in the absence of any medium or underlying mechanism because he is implicitly employing a different concept of causality, constant conjunction. Hence, he argues:

> If Mr. Spencer means that the action of gravitation gives us no sensations, the assertion is one than which I have not seen, in the writings of philosophers, many more startling. What other sensation do we need than the sensation of one body moving toward another? 'The elements of representation' are not two bodies and an 'agency', but two bodies and an effect; viz. the fact of their approaching one another. (J. S. Mill, *Collected Works*, ed. Robson, vol. VII, p. 275).

Since Mill does not think that the analysis of '*A* moves *B*' requires an agency, he is quite willing to describe a situation in which two bodies in isolation approach each other as a case of one body causing another

to move. Therefore, the disagreement between Mill and Spencer is over whether to describe a state of affairs consisting solely of two bodies moving toward each other as a case of one body causing another to move. They do not really disagree over the conceivability of the state of affairs in question.[1]

The University of Nebraska—Lincoln   © ALBERT CASULLO 1979

[1] An earlier version of this paper was presented at the Seventh Hume Conference, September 15–18, 1978, in Banff, Alberta, Canada.

QUEST ON THE ENTITLEMENT THEORY

By DAVID M. LEDERKRAMER

AN implication of Robert Nozick's entitlement theory of distributive justice (*Anarchy, State, and Utopia*, New York, 1974, Basic Books; all page references are to this book) is:

> N 'Whatever arises from a just situation [distribution] by just steps is itself just' (p. 151).

Edward Quest, in a paper titled after N (ANALYSIS 37.4), entertains three arguments which lead him to conclude that '... N appears to be false'. In the discussion which follows, I shall try to show that Quest's arguments are problematic and do not entitle him to this conclusion.

I

Quest begins in section I with the observation that 'N is not obvious . . .' In support of this observation he offers an argument which derives from his statement:

> Generally, if 'F' is inexact, it is not inconsistent to have a type of transfer which individually can never be pictured as changing an *x* which is F into an *x* which is not F even though a series of such transfers can be pictured as accomplishing this change (p. 205).

The implicit argument, a *modus ponens*, turns upon the assertion that the concept of distributive justice is 'inexact':

(1) If 'justice' is inexact, then some series of just exchanges is not justice-preserving.

(2) 'Justice' is inexact.

(3) Therefore, some series of just exchanges is not justice-preserving.

Because the concept of justice is inexact—that is, because the distinction between a just and an unjust pattern of holdings cannot be very finely drawn without being arbitrary—a small economic exchange will not upset the justness of a distribution but a large exchange will. Referring to Nozick's Wilt Chamberlain example Quest states

> Thus, the intuition that (4) a voluntary 25 cent exchange with Wilt Chamberlain could never be seen as transforming a just distribution into one which is not just . . . *is not sufficient for supporting* the belief that (5) any series of such exchanges will preserve justice: for if justice is inexact, a trivial exchange should not be expected to entitle anyone to make the non-trivial switch from 'the distribution is at present just' to 'now the distribution is not just' (p. 205, italics and parenthetical numbers mine).

Quest's criticism of Nozick is twofold: (a) N is insufficiently justified, because (4) above is not sufficient to support (5); and (b) N is false, on the ground that the conclusion of the 'inexactitude' argument—'Some series of just exchanges is not justice-preserving'—is the negation of a logically equivalent variant of N: 'Any series of just exchanges is justice-preserving'.

With respect to criticism (a), Quest, I suggest, has mistaken an illustration of N for its justification, and in so doing attacks an argument which Nozick does not propound. Quest claims (correctly) that (4) is not sufficient to support (5). He explains that (4), an obvious intuition, is not sufficient to support (5), a variant of N, by invoking the concept of inexactitude. But regardless of whether 'distributive justice' is an inexact concept in Quest's sense, (4) is not sufficient to support (5) on logical grounds alone. It cannot, without additional premises, be inferred from 'trivial exchanges are justice-preserving' that 'all exchanges are justice-preserving'. So it is not surprising that in justifying N Nozick does *not* argue that (4) is sufficient to support (5). Rather, Nozick employs a natural rights argument which begins with the supposition that any state of affairs is just if and only if no person's rights are violated. Accordingly, the criterion of a just distribution is that everyone has a (property) right to everything he possesses. Because property rights are exemplified by persons' entitlements to things, a distribution is just if everyone is entitled to his holdings. Since at the conclusion of a just transaction the parties to the transaction are entitled to their holdings, it follows that all just transactions are justice-preserving. Although this is a crude adumbration of Nozick's argument, it nevertheless can be seen that Quest's criticism is misdirected because N is justified on the basis of property rights, not upon the justice-preserving quality of trivial economic exchanges.

Criticism (b)—that the 'inexactitude' argument implies that N is

false—is not explicitly registered in Quest's paper, but in view of the subject under consideration it should not be overlooked. The 'inexactitude' argument derives from the distinction Quest draws between predicates which are inexact and those which are not. According to that distinction, inexact predicates are those whose extensions are ambiguously delimited. Consequently, a distinction drawn on the basis of an inexact predicate is susceptible to cases which do not fall clearly on either side of it. Numbered among inexact predicates are 'is heavy', 'is loud', and as Quest points out, 'is poor'. Further elucidating Quest's distinction, we can see why inexact predicates are inexact by noting the general rule for their ascription: 'x is F if x is $G > i$', where 'x' is an object, 'F' is an inexact predicate, 'G' is the property in virtue of which F is ascribed, and 'i' is the magnitude of G sufficient for something to be F. Thus, if x manifests G to a degree greater than i, then x is F. For example, 'x is heavy if x weighs more than 50 pounds'. The predicate is inexact in the sense that a single value for i cannot be specified without arbitrariness; or, to put the point differently, any number of values can be correctly substituted for i, but none in particular precisely marks the borderline between F and not-F.

We have, then, a criterion for distinguishing inexact predicates from others; inexact predicates are those which are ascribed according to the previously noted rule. The predicate 'is entitled to' is not, therefore, inexact. Thus, any theory of distributive justice which stipulates that 'a distribution is just if and only if everyone is entitled to his holdings' is not an inexact theory; that is, under such a theory the predicate 'is a just distribution' is not inexact. Because this is the case, premise (2) of the 'inexactitude' argument—'justice is inexact'—is not true of the entitlement theory, and, therefore, the argument is useless against Nozick.

II

The Chamberlain example hypothesizes a just distribution, D_1, under which a million people each choose to pay Wilt Chamberlain twenty-five cents in return for watching him play a game of basketball. Under D_2, the distribution at the end of the season, Chamberlain, having earned $250,000, is assumed to be the wealthiest man in the society. Nozick asks, R 'If D_1 was a just distribution, and people voluntarily moved from it to D_2 ... isn't D_2 also just?' (p. 161) In section II of his paper, Quest answers in the negative. He argues that it is false that people voluntarily moved from D_1 to D_2, and therefore that this cannot be the ground of D_2's justness. Quest states: '... it is possible that in voluntarily giving Chamberlain a quarter for the opportunity of watching a game, a spectator involuntarily becomes a participant in a social move, D_1 to D_2, unknowingly helping to bring about a distribution
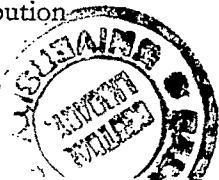
which that spectator detests.' In what follows I shall explain why this objection does not tell against the entitlement theory.

Suppose Quest is correct in arguing that the spectators did not voluntarily bring D_2 about. Does this objection undermine Nozick's position? I think not: Nozick can avoid Quest's criticism and still make his point by phrasing R as follows: 'If D_1 was a just distribution, and D_2 arose from the (just) transactions which people chose to conduct . . . isn't D_2 also just?' Thus, in spite of Quest's objection, Nozick's point—and his challenge to the reader—remains as follows: Under D_1— your favourite (just) distribution—each person is granted (at least) the right to possess and exchange a given amount of property. People choose to conduct certain non-fraudulent transactions, and D_2 results. The transactions which give rise to D_2 consist in the transfer of rights to possess and exchange certain property—rights granted under D_1. Thus, under D_2 everyone has acquired the right to possess what he holds. But the contention that D_2 is unjust implies that not everyone has the right to possess what he holds. Yet this was one of the very rights which was transferred. So Nozick's challenge—if you maintain that D_2 is unjust—is to explain how people have come to lose the rights which were transferred to them in the transition to D_2. If one accepts Nozick's initial assumptions—and Quest has not rejected them—one must either accept what they imply, that D_2 is just, or explain how people have come to lose the rights which they acquired through legitimate transactions. Quest fails to provide this explanation, and he doesn't argue against Nozick's initial assumptions, so why doesn't Nozick's conclusion stand? The requirement that people voluntarily choose the *distribution* is not a necessary step in Nozick's argument; that they voluntarily choose the *transactions* which account for the distribution is a necessary step, and it is satisfied by the Chamberlain example.

Thus, Quest's objection that the spectators did not voluntarily choose D_2 does not accomplish the task for which it was adduced. The objection is successful only if the entitlement theory is justified by, or implies, the claim that people should voluntarily choose distributions. But the entitlement theory is not justified by *that* claim, and it does not imply it—it implies only that if a distribution is just, people will have chosen the transactions which gave rise to it. Thus, as Quest claims, Nozick is incorrect in stating that the spectators voluntarily moved to D_2, but this does not reveal any inadequacy in the entitlement theory or in Nozick's defence of it.[1]

© DAVID M. LEDERKRAMER 1979

---

[1] I am grateful to the Editor for helpful criticism of an earlier draft.

IS RELATIVISM DISHONEST?

By David Gordon

SUPPOSE Alice accepts a moral theory according to which it is obligatory in ordinary circumstances to abstain from animal products. Then in appropriate circumstances she should advise or exhort Bob to abstain. She might believe, however, that there is an equally well grounded theory according to which not abstaining is permissible or obligatory. If Bob holds such a theory, and she advises him to abstain, she will be advising him to do something which in her opinion he has no reason, all things considered, to do. (Non-moral reasons are excluded from the example). Professor Postow argues (ANALYSIS 39.1, January, 1979) that she will thereby be dishonest.

His contention that Bob has no all-things-considered reason to abstain depends upon Bob's *adhering* to a well-grounded moral view denying that it is obligatory to abstain. The fact that such a view exists does not by itself make it dishonest for her to advise Bob to abstain. Otherwise, the fact that such a view exists would tell against *her* abstaining. Clearly, it does not, since she does not hold this view. Then, if Bob were to adopt Alice's view and abandon his own, no problem would occur. (Of course, all this assumes that it is consistent for Alice, without taking advice to other people into account, to be committed to a moral view even though she believes that an equally well grounded theory contradicts her own. Postow's argument assumes that this part of her position is consistent, and it is what he believes follows from this that I wish to challenge).

But Alice's advice to Bob just *is* an attempt to persuade him to change his opinion. If he accepts her advice, he will have come to believe that it is obligatory to abstain from animal products, which by the terms of the example is a well-grounded moral theory. A counter that suggests itself is that he might abstain while continuing to believe his previous position or that he might profess both views at once, in which cases he has no all-things-considered reason to abstain. But then the problem is an incoherence in his position, not in Alice's.

Postow might reply that if Bob adopts Alice's position, then there is no problem for him; but why should Alice induce him to abandon a position as well grounded as her own? To this the answer is simply: why shouldn't she? On Postow's position, she is required to try, and she is not urging him to do something which, if he listens to her, he will have no all-things-considered reason to do.

STEINER ON CARTESIAN SCEPTICISM

By DAVID GORDON

MARK Steiner, in 'Cartesian Scepticism and Epistemic Logic' (ANALYSIS 39.1, January, 1979), argues that even if the principle $K(S) \rightarrow KK(S)$ is rejected, weaker and plausible principles allow a version of Descartes' argument from scepticism based on dreams to go through. His reasons for questioning Hintikka's principle are derived from the Causal Theory of Knowledge. It seems to me that his weaker principles fall victim to the same sort of challenge as he raises to Hintikka's principle.

The stronger of his two substitute principles is: 'If it is irrational to assert $K(P)$, then it is irrational to assert P.' (p. 40). This seems inconsistent with his example from the Causal Theory, in which a student correctly solves a problem but has little confidence in his results. Although the student's lack of confidence in his results makes it irrational for him to claim that he knows that his results are correct, it is not irrational for him to answer.

The weaker principle 'If one is committed to $-K(P)$, then it is irrational of him to assert P' also seems dubious. Suppose the student is quite sure that he lacks confidence in his results; it still does not follow that it is irrational for him to answer the test question. It seems to me that Professor Steiner has failed to show that modifications of Hintikka's principle can be produced which are consistent with the Causal Theory while allowing a version of the Cartesian argument to stand.

© DAVID GORDON 1979